Humboldt-Universität zu Berlin

Numerical methods for backward stochastic differential equations of quadratic and locally Lipschitz type

Dissertation

zur Erlangung des Grades:

Dr. rer. nat. im Fach Mathematik

an der Mathematisch-Naturwissenschaftlichen Fakultät II

vorgelegt von

MMATH, Plamen Turkedjiev

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät II:
Prof. Dr. Elmar Kulke

Datum der Verteidigung: 30. April 2013

Vorsitzende der Promotionskommission: Prof. Dr. Peter Imkeller

Gutachter: Prof. Dr. Dirk Becherer, Prof. Dr. Emmanuel Gobet,
Prof. Dr. John Schoenmakers

Acknowledgments

As anyone whose done a PhD knows very well, writing the thesis is not a solitary undertaking. There is an enormous list of people that help, directly or indirectly, and I would like to take a moment to thank an extremely in-exhaustive portion of those people. To those I miss out, I am deeply sorry, know only that I am greatly in your debt.

First of all, I would like to thank my supervisor, Prof. Dirk Becherer, for giving me the opportunity to study this wonderful topic and for shepherding me through the arduous process of completing my thesis and learning the process of research. I am very grateful for our many mathematical discussions, which have served to substantially improve the quality of the work and to generate many ideas, and also for his excellent career advise. He is an excellent mathematician and mentor, and a resourceful collaborator, and a very kind person.

Secondly, I would like to thank Prof. Emmanuel Gobet. During several research visits, I profited greatly from his immense knowledge, creativity, energy, and kindness. His input has had an invaluable impact on this thesis. It was always a pleasure to do research with him, and I am very honoured to be able to continue working with him in my first postdoctoral position at École Polytechnique.

Thirdly, I'd like to thank the entire viva committee for their meticulous attention to detail in the examination of the thesis. The final version thesis was substantially improved thanks to their comments and corrections.

Fourthly, I would like to thank friends, colleagues and loved-ones for making life outside work so joyful and fulfilling.

Finally, I would like to thank my wonderful parents, who have always been supporting and encouraging. Your pride and love has enriched me.

I was supported financially, at different stages, by the IRTG "Stochastic Models of Complex Processes", DFG Research Center MATHEON, and Humboldt Universität zu Berlin.

Abstract

The focus of the thesis is to develop efficient numerical schemes for quadratic and locally Lipschitz decoupled forward-backward stochastic differential equations (BSDEs). The terminal conditions satisfy weak regularity conditions. Although BSDEs have valuable applications in the theory of financial mathematics, stochastic control and partial differential equations, few efficient numerical schemes are available. Three algorithms based on Monte Carlo simulation are developed. Starting from a discrete time scheme, least-square regression is used to approximate conditional expectation. One benefit of these schemes is that they require as an input only the simulations of an explanatory process X and a Brownian motion W. Due to the use of distribution-free tools, one requires only very weak conditions on the explanatory process X, meaning that these methods can be applied to very general probability spaces. Explicit upper bounds for the error are obtained. The algorithms are then calibrated systematically based on the upper bounds of the error and the complexity is computed. Using a time-local truncation of the BSDE driver, the quadratic BSDE is reduced to a locally Lipschitz BSDE, and it is shown that the complexity of the algorithms for the locally Lipschitz BSDE is the same as that of the algorithm of a uniformly Lipschitz BSDE. It also is shown that these algorithms are competitive compared to other available algorithms for uniformly Lipschitz BSDEs.

In the first chapter, the key issue of the global error due to discrete time approximation is handled in the Brownian setting. A priori estimates for the martigale integrand Z are obtained using a representation theorem. These estimates are the basis of the error analysis - there is no use of BMO techniques. Using non-uniform time-grids, whose time-points are more concentrated at the end of the time interval, one is able to obtain comparable convergence rates as in the uniform Lipschitz driver setting, and distinct conditions are given under which the convergence rate is weaker. Then, a novel discrete time approximation of the BSDE using Malliavin weights is constructed and the approximation error is analyzed. In the second chapter, a numerical algorithm based on multistep forward dynamical programming is developed. This is similar to the Bender-Denk scheme, except that there are no Picard iterations and the theory is extended to the locally Lipschitz continuous BSDEs. The lack of Picard iterations also allows a refinement of the complexity analysis. In the third chapter, variance reduction techniques are used to improve the efficiency of the algorithm of the second chapter. The BSDE is decomposed into a system of a linear BSDE and a nonlinear BSDE with zero terminal condition. It is shown that it more efficient to solve the system numerically than to solve the original BSDE numerically. The main focus of the chapter is to develop a multilevel algorithm for the approximation of the linear BSDE. The use of multilevel is novel in the context of BSDEs. The efficiency of the multilevel algorithm is higher than that of algorithm of the second chapter, leading to further improvement in the overall efficiency of the BSDE approximation. However, the validity of the results for the multilevel scheme are restricted to the uniformly Lipschitz continuous terminal condition and driver, and it is restricted to the probability space generated by the Brownian motion. In the fourth and final chapter, a novel algorithm based on the Malliavin weights discrete time approximation from the first chapter scheme is developed and analyzed. It proves to be more efficient than the algorithm of the second chapter and is not restricted to the Lipschitz continuous terminal condition.

Zussamenfassung

Der Fokus dieser Dissertation liegt darauf, effiziente numerische Methode für ungekoppelte lokal Lipschitz-stetige und quadratische stochastische Vorwärts-Rückwärtsdifferenzialgleichungen (BSDE) mit Endbedingungen von schwacher Regularität zu entwickeln. Obwohl BSDE viele Anwendungen in der Theorie der Finanzmathematik, der stochastischen Kontrolle und der partiellen Differenzialgleichungen haben, gibt es bisher nur wenige numerische Methoden. Drei neue auf Monte-Carlo-Simulationen basierende Algorithmen werden entwickelt. Die in der zeitdiskreten Approximation zu lösenden bedingten Erwartungen werden mittels der Methode der kleinsten Quadrate näherungsweise berechnet. Ein Vorteil dieser Algorithmen ist, dass sie als Eingabe nur Simulationen eines Vorwärts-
prozesses X und der Brownschen Bewegung benötigen. Da sie auf modellfreien Abschätzungen aufbauen, benötigen die hier vorgestellten Verfahren nur sehr schwache Bedingungen an den Prozess X. Daher können sie auf sehr allgemeinen Wahrscheinlichkeitsräumen angewendet werden. Für die drei numerischen Algorithmen werden explizite maximale Fehlerabschätzungen berechnet. Die Algorithmen werden dann auf Basis dieser maximalen Fehler kalibriert und die Komplexität der Algorithmen wird berechnet. Mithilfe einer zeitlich lokalen Abschneidung des Treibers der BSDE werden quadratische BSDE auf lokal Lipschitz-stetige BSDE zurückgeführt. Es wird gezeigt, dass die Komplexität der Algorithmen im lokal Lipschitz-stetigen Fall vergleichbar zu ihrer Komplexität im global Lipschitz-stetigen Fall ist. Es wird auch gezeigt, dass der Vergleich mit bereits für Lipschitz-stetige BSDE existierenden Methoden für die hier vorgestellten Algorithmen positiv ausfällt.

Im ersten Kapitel werden globale Fehlerabschätzungen für zeitdiskrete Approximationen von BSDE in Brownschen Kontext behandelt. Basierend auf einem Darstellungssatz für den Martingal Integranden Z werden neue a priori Abschätzungen hergeleitet. Die a priori Abschätzungen bilden die Grundlage der Fehleranalyse und BMO Methoden werden nicht verwendet. Mithilfe ungleichmäßiger Zeitgitter, deren Zeitpunkte am Ende des betrachteten Zeitintervalls dichter sind, wird eine vergleichbare Konvergenzrate wie im Lipschitz-stetigen Fall erreicht, und es werden explizite Bedingungen dafür angegeben, wann die Konvergenzrate schlechter ist. Zum Schluss des Kapitels wird eine neuartige zeitdiskrete Approximation der BSDE entwickelt, die auf der Verwendung von Malliavin-Gewichten basiert, und die Fehlerabschätzung für diese neue Approximation wird berechnet. Im zweiten Kapitel wird ein auf der "multistep forward" dynamischen Programmierung basierendes numerisches Verfahren entwickelt. Dieses Verfahren weist Ähnlichkeiten zum Bender-Denk Algorithmus auf aber es benötigt keine Picard-Iterationen. Da keine Picard-Iterationen verwendet werden, lässt sich die Komplexitätsanalyse verfeinern. Im dritten Kapitel werden Varianzreduktionsmethoden angewendet, um die Effizienz des im zweiten Kapitel entwickelten Algorithmus zu verbessern. Die BSDE wird in ein aus einer linearen BSDE mit Null-Treiber und einer nichtlineare BSDE mit Null-Endbedingung bestehendes System zerlegt. Es wird gezeigt, dass es effizienter ist, das System von BSDE numerisch zu lösen als die ursprüngliche BSDE numerisch zu lösen. Der Hauptfokus des Kapitels liegt darauf, eine Multilevel Methode zur Approximation von BSDE zu entwickeln. Im Kontext von BSDE wurden Multilevel Methoden bisher nicht verwendet. Der Multilevel Algorithmus ist effizienter als der im zweiten Kapitel vorgestellte Algorithmus, was zu einer weiteren Verbesserung der Effizienz der Approximation von BSDE führt. Die Methoden dieses Kapitels beschränken sich auf BSDE mit Lipschitz-stetigen Endbedingungen und Treibern sowie auf den von der Brownschen Bewegung erzeugten Wahrscheinlichkeitsraum. Im vierten Kapitel wird ein neuer Algorithmus entwickelt, der auf der im ersten Kapitel vorgestellten, auf Malliavin-Gewichten basierenden, zeitdiskreten Approximation aufbaut. Es wird gezeigt, dass dieser Algorithmus numerisch effizienter ist als der im zweiten Kapitel vorgestellte Algorithmus.

Contents

1 Discrete time approximations of locally Lipschitz and quadratic backward stochastic differential equations

1.1 Introduction

Let $T > 0$ be a fixed terminal time and $(\Omega, \mathcal{F}_T, \{\mathcal{F}_t\}, \mathbb{P})$ a filtered probability space, where $\{\mathcal{F}_t : 0 \leq t \leq T\}$ is the filtration generated by a q-dimensional ($q \geq 1$) Brownian motion W and satisfying the usual conditions of right-continuity and completeness. The focus of this chapter will be on the approximation by discrete time processes of the pair of $\mathbb{R} \times (\mathbb{R}^q)^\top$-valued, where $(\mathbb{R}^q)^\top$ is the space of q-dimensional real valued row vectors, predictable processes (Y, Z) that solve a backward stochastic differential equation (BSDE) of the form

$$Y_t = \Phi(X_T) + \int_t^T f(s, X_s, Y_s, Z_s)ds - \int_t^T Z_s dW_s. \tag{1.1.1}$$

where $\Phi : \mathbb{R}^d \to \mathbb{R}$ and $f : [0, T) \times \mathbb{R}^d \times \mathbb{R} \times (\mathbb{R}^q)^\top \to \mathbb{R}$, and X is the solution of the $\mathbb{R}^d$-valued ($d \leq q$) stochastic differential equation (SDE)

$$X_t = x_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s, \tag{1.1.2}$$

where $b : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : [0, T] \times \mathbb{R}^d \to \mathbb{R}^{d\times q}$, and x_0 is a fixed vector in $\mathbb{R}^d$.

BSDEs have a myriad of applications, notably in the theory of mathematical finance, stochastic control, and partial differential equations. Although the theory and applications of BSDEs has grown rapidly over the last twenty years, the development of efficient numerical methods is a more recent topic of research.

The existence and uniqueness of the BSDE (1.1.1) is highly dependent on coefficients b and σ of the SDE and on the terminal condition Φ and the driver f of the BSDE, as is the quality of the numerical schemes used to approximate them. The majority of work in the field of numerical methods for BSDEs has been in the setting where the terminal condition Φ and the driver f are Lipschitz continuous in (x, y, z) uniformly in t. A good overview can be found in [GL10]. More recently, several authors have started to develop numerical methods to handle BSDEs satisfying weaker conditions. An important extension has been to quadratic BSDEs, where the driver f is allowed to satisfy a quadratic growth in z: [IDR10] considered the setting where the terminal condition Φ is bounded and Lipschitz continuous, and [Ric11] considers the setting in which Φ is bounded but only Hölder continuous. The case of the BSDE with quadratic driver f is particularly interesting in financial mathematics. It has applications in, for example, utility optimization in incomplete markets [REK00][HIM05]. On the other hand [GM10] considered the setting in which driver f is globally Lipschitz continuous but the terminal condition Φ is allowed to satisfy very weak regularity conditions. [GGG12] extend the work of [GM10] in that they consider terminal conditions that depend on finitely many points along the path of X, so $\Phi(X_T)$ is replaced by $\Phi(X_{t_1}, \dots, X_{t_L})$ in (1.1.1) above.

In this chapter, the framework of [GM10] is extended to the BSDE where the driver is **locally Lipschitz continuous in** (x, y, z) and **locally bounded at** 0. To be precise, let $f : [0, T] \times$

$\mathbb{R}^d \times \mathbb{R} \times (\mathbb{R}^q)^\top \to \mathbb{R}$ be continuous and assume there exist constants $\theta_L,\ \theta_c \in (0,1]$ and finite $L_f, C_f \geq 0$ such that, for all $t \in [0,T)$ and $(x,y,z),(x',y',z') \in \mathbb{R}^d \times \mathbb{R} \times (\mathbb{R}^q)^\top$,

$$|f(t,x,y,z) - f(t,x',y',z')| \leq L_f \frac{|x-x'| + |y-y'| + |z-z'|}{(T-t)^{(1-\theta_L)/2}}, \quad |f(t,x,0,0)| \quad \leq \frac{C_f}{(T-t)^{1-\theta_c}} \tag{1.1.3}$$

The coefficients of the SDE b and σ are bounded, twice continuously differentiable in x, with bounded and Hölder continuous derivatives, and $\frac{1}{2}$-Hölder continuous in t. Additionally, σ is uniformly elliptic. We will work with terminal conditions Φ in the function space

$$\mathbf{L}_{2,\alpha} := \Big\{ g : \mathbb{R}^d \to \mathbb{R} \ : \ \mathbb{E}[g(X_T)^2] + \sup_{t \in [0,T)} \frac{\mathbb{E}[|g(X_T) - \mathbb{E}[g(X_T)|\mathcal{F}_t]|^2]}{(T-t)^\alpha} < \infty \Big\}.$$

for some regularity parameter $\alpha \in (0,1]$. It is clear that $\cup_{\alpha \in (0,1]} L_{2,\alpha}$ contains the Hölder continuous functions. This space is rather natural for the generalization beyond classically Lipschitz and Hölder continuous terminal conditions, and it has been studied in [GM10] and [GG11], amongst others.

The class of locally Lipschitz continuous BSDEs with terminal conditions in $\bigcup_{\alpha \in (0,1]} \mathbf{L}_{2,\alpha}$ is a very large class of BSDEs. Of course, the usual BSDEs with uniformly Lipschitz continuous terminal condition and driver are contained in this class. Moreover, it turns out that if (Y, Z) satisfy the BSDE (1.1.1) with quadratic driver f (under some usual conditions) and bounded, Hölder continuous terminal condition Φ, it is possible to apply a time local truncation of the driver that reduces it to a locally Lipschitz driver of the form (1.1.3) without affecting the solution of the BSDE; see Section 1.3.1. Therefore, a successful numerical method for locally Lipschitz BSDEs will also work well for quadratic BSDEs.

At this point, it is useful to give a brief comparison between the assumptions of this chapter and those of [Ric11]. Firstly, whilst this chapter requires a uniformly elliptic condition on σ, this condition is relaxed in [Ric11]. Secondly, the coefficient $\sigma = \sigma(t)$ is only allowed to depend on t in [Ric11], whereas we are able to incorporate a dependence on x.

An important result of Chapter 1, Theorem 1.5.1, is not necessarily associated to numerical schemes, but will be used extensively and is a corner stone of our analysis. We prove a representation theorem, showing that Z has a version $\mathcal{Z}$ which satisfies

$$\mathcal{Z}_t = \mathbb{E}_t[\Phi(X_T) H_T^t + \int_t^T f(s, X_s, Y_s, Z_s) H_s^t ds] \quad \text{for all } t \in [0,T) \text{ almost surely,}$$

where the processes H_r^t are the so called Malliavin weights

$$H_r^s = \frac{1}{r-s} \Big(\int_s^r (\sigma^{-1}(t, X_t) D_s X_t)^\top dW_t \Big)^\top$$

where $\sigma^{-1}(\cdot)$ is a right-inverse of the matrix $\sigma(\cdot)$, and $D_\cdot X_t$ is the Malliavin derivative of X_t. This is an extension of the results of [MZ02, Theorem 4.2], who work in the setting where Φ and f are uniformly Lipschitz continuous. Rather conveniently, the representation of Z_t does not require the derivatives of Φ and f, which will allow us to take advantage of the fact that Φ is in the space $\mathbf{L}_{2,\alpha}$ and the the driver f is locally Lipschitz continuous and bounded as in (1.1.3) but not differentiable. The representation (1.5.1) leads to stability results - so called a priori estimates - on the difference of two BSDEs $\mathbb{E}[|Z_t - \bar{Z}_t|^2]$ in Proposition 1.4.2. This is in addition to the usual time-averaged a priori estimates on $\int_0^T \mathbb{E}[|Z_t - \bar{Z}_t|^2]dt$ as in, for example, [EKPQ97, Proposition 2.1]. This leads to (Corollary 1.4.4) the bounds $\mathbb{E}[|Z_t|^2] \leq C(T-t)^{\left((2\theta_c \wedge \alpha - 1\right)/2}$, for a time t independent constant C, when the terminal condition Φ is in $\mathbf{L}_{2,\alpha}$. A generalization of the a priori estimates using conditional expectations in Lemma 1.4.5 allows the partial calibration of the time local truncation of the quadratic driver in Remark 1.4.7. Namely, if the terminal condition Φ is bounded and α-Hölder continuous, then the time local truncation reduces the quadratic driver to a locally Lipschitz continuous driver with $\theta_L = \alpha$.

The first step to finding a good numerical method for the BSDE (1.1.1) is to find a suitable discrete-time process for the approximation. This means setting a time-grid $\pi = \{0 = t_0 < \cdots < t_N = T\}$ with $N+1$ time points and finding piecewise constant adapted processes (Y^π, Z^π) given by

$$Y_t^\pi := \sum_{i=0}^{N-1} Y_i^\pi \mathbf{1}_{[t_i, t_{i+1})}(t), \qquad Z_t^\pi = \sum_{i=0}^{N-1} Z_i^\pi \mathbf{1}_{(t_i, t_{i+1}]}(t)$$

where Y_i^π and Z_i^π are $\mathcal{F}_{t_i}$-measurable random variables. This process consists of two parts: first, finding a suitable time-grid, and second, computing the random variables $(Y_i^\pi, Z_i^\pi)_{0 \leq i < N}$. If Y and Z are in suitable square integrable spaces, the natural measure for the error of the discrete approximation is

$$\max_{0 \leq i \leq N-1} \mathbb{E}[|Y_{t_i} - Y_i^\pi|^2] + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathbb{E}[|Z_t - Z_i^\pi|^2]dt. \tag{1.1.4}$$

A rather natural approximation (Y^π, Z^π) is based on the so called One-step Dynamical Programming (ODP) equation:

$$\begin{aligned} Y_N = \Phi(X_T), \quad Z_i &= \frac{1}{t_{i+1} - t_i} \mathbb{E}[Y_{i+1}(W_{t_{i+1}} - W_{t_i})^\top | \mathcal{F}_{t_i}], \\ Y_i &= \mathbb{E}[Y_{i+1} + f(t_i, X_{t_i}, Y_{i+1}, Z_i)(t_{i+1} - t_i) | \mathcal{F}_{t_i}] \end{aligned} \tag{1.1.5}$$

for $i \in \{0, \ldots, N-1\}$. The ODP and its implementation has been frequently studied [Zha04][BT04] [LGW06]. We shall develop numerical methods for BSDEs based on the ODP time discrete approximation (1.1.5) in Chapters 2 and 3. In the case where the driver f is uniformly Lipschitz, one can show using the methods of [GL06, Theorem 1] that the error of this scheme (1.1.4) is bounded above by $C|\pi| + C\mathcal{E}(\pi)$ for some constant C, that does not depend on the time-grid π (or the number of points in it N), and the values

$$|\pi| = \max_{0 \leq i \leq N-1} (t_{i+1} - t_i), \quad \bar{Z}_i := \frac{1}{t_{i+1} - t_i} \mathbb{E}[\int_{t_i}^{t_{i+1}} Z_t dt | \mathcal{F}_{t_i}], \qquad \mathcal{E}(\pi) := \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathbb{E}[|Z_t - \bar{Z}_i|^2]. \tag{1.1.6}$$

The quantity $\mathcal{E}(\pi)$ is often called the $\mathbf{L}_2$-regularity. In fact, in [GL06], one assumes that the terminal condition be Lipschitz continuous, but in fact this does not play a role in our formulation of the ODP (1.1.5) because we are not approximating the SDE X. If we wish to incorporate discrete time approximations $(X_i)_i$ of X, we require that it satisfies the conditions

$$\mathbb{E}[|\Phi(X_T) - \Phi(X_N)|^2] \leq C_X|\pi|, \qquad \mathbb{E}[|X_{t_i} - X_i|^2] \leq C_X|\pi| \quad \forall i \in \{0, \ldots, N\}$$

for a constant C_X that is independent of N and the size of the time increments of π. Discrete time approximations of SDEs are not the topic of this thesis, rather we assume that such a suitable approximation is available. The book of [KP92] is a good reference for the approximation of SDEs. It is also easy to generalize [GL06, Theorem 1] to the locally Lipschitz driver satisfying (1.1.3): one must simply select the time increments sufficiently small to compensate for the dependence on time in the Lipschitz coefficient $L_f(T-t_i)^{(\theta_L-1)/2}$, but otherwise proceed as for the uniformly Lipschitz case. Therefore, it is useful to obtain an upper bound for $\mathcal{E}(\pi)$ to determine an upper bound for the error of the ODP scheme (1.1.5). In fact, determining an upper bound for $\mathcal{E}(\pi)$ that converges to 0 as $|\pi| \to 0$ is a challenging task that has been studied in many works. For example, it was shown in the pioneering work of [Zha04, Theorem 3.1] that $\mathcal{E}(\pi) \leq C|\pi|$ in the case of the uniformly Lipschitz continuous terminal Φ condition and driver. In [GM10], the authors extend the results of [Zha04, Theorem 3.1] to the case where the terminal condition Φ is a function in $\mathbf{L}_{2,\alpha}$ and the driver f is uniformly Lipschitz continuous. They make use of a special class of time-grid

$$\left\{\pi_N^{(\beta)} = \{t_i = T - T(1-i/N)^{1/\beta} \ : \ i = 0, \ldots, N\} \ : \ \beta \in (0,1]\right\}$$

and demonstrate [GM10, Theorem 3.3] that for $\Phi \in \mathbf{L}_{2,\alpha}$ and $\beta < \alpha$, there is a constant C_β, which depends on β but not on N, such that $\mathcal{E}(\pi_N^{(\beta)}) \leq C_\beta N^{-1}$. This time-grid appears to be optimal: using the uniform time-grid $\pi_N = \{t_i = Ti/N : i = 0, \ldots, N\}$ one can only obtain $\mathcal{E}(\pi_N) \leq CN^{-\alpha}$ for some constant C independent of N. In fact, it is not even possible to obtain the optimal bound with $\pi_N^{(\alpha)}$ [GM10, Section 1.2].

In this chapter, the results of [GM10] are extended to the locally Lipschitz and locally bounded driver f satisfying (1.1.3). This implies that we are also treating the quadratic BSDE with bounded, Hölder continuous terminal condition. The time-grids $\pi_N^{(\beta)}$ again play an important role. Unlike in the case of uniformly Lipschitz continuous driver f, we are not able to obtain the optimal rate of convergence N^{-1} of $\mathcal{E}(\pi_N^{(\beta)})$ for all α, θ_L and θ_c. By first of all applying the techniques of [GM10] directly, we show in Proposition 1.6.2 that if $\beta < (2\gamma) \wedge \alpha$,

$$\mathcal{E}(\pi_N^{(\beta)}) \leq CN^{-1} + CN^{(1-(\alpha+\theta_L)\wedge 1)/\gamma - 1}$$

where $\gamma = (\frac{\alpha}{2} \wedge \theta_c + \frac{\theta_L}{2}) \wedge \theta_c$. The constant C does not depend on N, but may depend on β. The optimal rate of convergence N^{-1} is obtained if $\alpha + \theta_L \geq 1$. In the special case where $\theta_L = \alpha$, which is the case of interest for the quadratic BSDE, this implies that the optimal rate N^{-1} is obtained for $\alpha \geq 1/2$. These conditions can be extended with the help of the a priori estimates and the additional assumption that the terminal condition has exponential moments. In Theorem 1.6.6, we show that if $\beta < (2\gamma) \wedge \alpha$, then

$$\mathcal{E}(\pi_N^{(\beta)}) \leq CN^{-1} + CN^{-2+(3\theta_L/4-1)(1+\delta_N)/(2\gamma)}(\ln(N) \vee 1)$$

where $\delta_N \geq \mathbf{1}_{[3,\infty)}(N) \ln\ln(N)/\ln(N)$. The constant C does not depend on N. Now $7\gamma + 2\theta_L \geq 4$

is sufficient to obtain the optimal convergence rate N^{-1}. In the special case $\alpha = \theta_L$ and $\theta_c = 1$, which is associated with the quadratic BSDE, this implies that the optimal rate of convergence is obtained for $\alpha \geq 4/9$, which is an improvement of the result from Proposition 1.6.2. The results of Theorem 1.6.6 require higher integrability conditions on the terminal condition and may impose more constraints on the parameter θ_c than Proposition 1.6.2, so it is useful to have both the results of Proposition 1.6.2 and Theorem 1.6.6.

It is interesting to observe that, just as for the uniformly Lipschitz continuous driver f, the use of the time-grids $\pi^{(\beta)}$ appears to substantially improve the numerical resolution of the locally Lipschitz BSDE, in the sense that a better rate of convergence is obtained for $\mathcal{E}(\pi)$. A rather different, but also non-uniform, time-grid was used in [Ric11, Theorem 4.17] to study quadratic BSDEs. In that work, the author is able to report an error bound $C_\eta N^{-\alpha+\eta}$, for any $\eta > 0$ and constant C_η depending possibly on η but not on N, for his approximation scheme.

Our techniques are dependent on a priori estimates derived from the point-wise representation (1.5.1) of Z_t. There is no use of BMO estimates. This means that it may be possible to extend these methods to study the numerical resolution of quadratic BSDEs with unbounded terminal condition [BH08]. Since [BH08] require the terminal condition to have exponential moments, the results of Theorem 1.6.6 will apply in this context. Moreover, our methods do not require special smoothing of the terminal condition and seem quite capable of handling functions in $\bigcup_{\alpha\in(0,1]} \mathbf{L}_{2,\alpha}$. It may be possible to extend these methods to the degenerate setting: indeed, [Zha05, Theorem 5.2] gives a representation theorem in just such a setting.

In the final part of this chapter, we propose a novel discrete-time scheme based on the representation with Malliavin weights of Z_t. This takes the form

$$Y_i := \mathbb{E}[\Phi(X_N) + \sum_{j=i+1}^{N-1} f_j(X_j, Y_j, Z_j)(t_{j+1} - t_j)|\mathcal{F}_{t_i}], \tag{1.1.7}$$

$$Z_i := \mathbb{E}[\Phi(X_N)H_N^i + \sum_{j=i+1}^{N-1} f_j(X_j, Y_j, Z_j)H_j^i(t_{j+1} - t_j)|\mathcal{F}_{t_i}] \tag{1.1.8}$$

for $i \in \{0, \ldots, N-1\}$, where $(X_i)_i$ is a suitable discrete time approximation of the SDE X and $(H_j^i)_{i,j}$ is a suitable discrete approximation of the Malliavin weights $(H_r^s)_{s,r}$. It turns out that the error of the approximation (1.1.4) cannot be controlled by the $\mathbf{L}_2$-regularity $\mathcal{E}(\pi)$ alone, as for the ODP scheme (1.1.5), but additionally requires control of the more complicated terms

$$\Big(\sum_{j=0}^{N-1}\frac{\int_{t_j}^{t_{j+1}}\{\|Y_r - Y_{t_j}\|_2 + \|Z_r - Z_{t_j}\|_2\}dr}{(T-t_j)^{(1-\theta_L)/2}}\Big)^2 + \sum_{i=0}^{N-2}\Big(\sum_{j=i+1}^{N-1}\frac{\int_{t_j}^{t_{j+1}}\{\|Y_r - Y_{t_j}\|_2 + \|Z_r - Z_{t_j}\|_2\}dr}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}}\Big)^2(t_{i+1} - t_i).$$

We use again the time-grids $\pi_N^{(\beta)}$, and compute in Theorem 1.7.6 that, if $\beta < (2\gamma) \wedge \alpha \wedge \theta_L$, then the error of the Malliavin weights scheme is bounded above by

$$CN^{-1} + C\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} \mathbb{E}[|Z_t - Z_{t_i}|^2]dt + C\big(N^{-2\gamma} + N^{(1-3\theta_L/4)/(2\gamma)-2}\big)(\ln(N)\vee 1)$$

where $\gamma = (\frac{\alpha}{2}\wedge\gamma + \frac{\theta_L}{2})\wedge\theta_c$. This is discussed in Remark 1.7.7. Exponential bounds of the terminal condition are essential for this result. The constant C does not depend on N, but it may depend on β. In the special case where $\alpha = \theta_L$ and $\theta_c = 1$, which includes the quadratic BSDE, this implies that the optimal rate of convergence CN^{-1} of the approximation error is obtained only for α greater than $1/2$. Therefore, the optimal rate of convergence achieved for fewer values of α than for the ODP scheme.

The reason that we are interested in the scheme with Malliavin weights (1.1.8) is that it appears to be quite suitable for computational purposes. In fact, it seems that this scheme may be more efficient on the numerical, rather than time-discretization level, than the ODP scheme 1.1.5 when the conditional expectations are approximated by Monte Carlo least-squares regression. The algorithm based on the Malliavin weights scheme is investigated in great detail in Chapter 4.

1.1.1 Notation and conventions

Let n, k, and l be non-zero integers whose value may change depending on the context.

The conditional expectation $\mathbb{E}[\cdot|\mathcal{F}_t]$ is denoted by $\mathbb{E}_t[\cdot]$ and the norm $\sqrt{\mathbb{E}[|\cdot|^2]}$ by $\|\cdot\|_2$.

For any Euclidean space E, $\mathcal{B}(E)$ denotes the Borel measurable sets in E, and the Lebesgue measure on the measurable space $\big(E,\mathcal{B}(E)\big)$ is denoted by m. For a $\mathcal{B}(E)$-measurable function f, the integral $m(f)$ of f with respect to the Lebesgue measure is denoted by $\int_E f(x)dx$. $\mathbf{L}_2(E;(\mathbb{R}^k)^\top)$ denotes the space of $\mathcal{B}(E)$-measurable functions $f : E \to (\mathbb{R}^k)^\top$ such that $\int_E |f(t)|^2 dx$ is finite.

$\mathbf{L}_2(\mathcal{F}_T)$ is the space of random variables that are square integrable with respect to $\mathbb{E}$. For any sub-σ-algebra $\mathcal{G}\subset\mathcal{F}_T$, $\mathbf{L}_2(\mathcal{G})\subset\mathbf{L}_2(\mathcal{F}_T)$ is the space of square integrable $\mathcal{G}$-measurable random variables.

The measure $m\times\mathbb{P}$ on the measurable space $\big([0,T)\times\Omega,\mathcal{B}([0,T))\otimes\mathcal{F}_T\big)$ denotes the product measure of m and $\mathbb{P}$. $\mathbf{L}_2([0,T]\times\Omega)$ is the space of $\mathcal{B}\big([0,T)\big)\otimes\mathcal{F}_T$-measurable processes that are square integrable with respect to $m\times\mathbb{P}$. $\mathbf{L}_2([0,T)\times\Omega;\mathbb{R}^k)$ denotes the processes in $\mathbf{L}_2([0,T)\times\Omega)$ taking values in $\mathbb{R}^k$. For two processes X and Y in $\mathbf{L}([0T)\times\Omega;\mathbb{R}^k)$, Y is said to be a version of X if $X = Y$ $m\times\mathbb{P}$-a.e. $\mathcal{P}\subset\mathcal{B}([0,T])\otimes\mathcal{F}_T$ is the predictable σ-algebra, generated by the continuous, adapted processes, and $\mathcal{H}^2$ is the subspace of $\mathbf{L}_2([0,T]\times\Omega)$ containing only predictable processes. For $p\geq 2$, $\mathcal{S}^p$ is the subspace of $\mathcal{H}^2$ of continuous processes Y such that $\|Y\|_{\mathcal{S}^p} := \mathbb{E}[\sup_{0\leq s\leq T}|Y_s|^p]^{\frac{1}{p}}$ is finite for all $Y\in\mathcal{S}^p$; $\|\cdot\|_{\mathcal{S}^p}$ is a norm for this space.

We will consider terminal conditions of the form $\Phi(X_T)$, where $x\mapsto\Phi(x)$ is in the space of fractionally smooth functions, $\mathbf{L}_{2,\alpha}$. To be precise, $\Phi:\mathbb{R}^d\to\mathbb{R}$ is a measurable function and there exists a constant $\alpha \in (0,1]$ such that $K^\alpha(\Phi) < \infty$, where

$$\left.\begin{aligned} K^\alpha(\Phi)^2 &:= \mathbb{E}[|\Phi(X_T)|^2] + \sup_{0\leq t<T} \tfrac{V_{t,T}(\Phi)^2}{(T-t)^\alpha} \\ \text{for } V_{t,T}(\Phi)^2 &:= \mathbb{E}[|\Phi(X_T) - \mathbb{E}_t[\Phi(X_T)]|^2]. \end{aligned}\right\} \tag{1.1.9}$$

We remark that any α-Hölder continuous function g belongs to $\mathbf{L}_{2,\alpha}$. In fact, it may be the case that the fractional smoothness of an α-Hölder continuous function g is better than α, i.e. that there there is a $\beta \in (\alpha, 1]$ such that $g \in \mathbf{L}_{2,\beta}$; an explicit example is given in [GGG12, Section 2, page 2087] for examples. We refer to [GM10] for further discussion of and references for the space $\mathbf{L}_{2,\alpha}$.

We identify the space of $k \times n$ dimensional, real valued matrices with $\mathbb{R}^{k\times n}$. For any vector or matrix A, $A^\top$ denotes its transpose. For any two vector V and U, the dot product $V^\top U$ is denoted by $V \cdot U$. In stochastic integrals where the integrator V is $\mathbb{R}^k$ valued and integrand U is $(\mathbb{R}^k)^\top$-valued, we write $\int U_t dV_t$ to mean $\sum_{i=1}^k \int U_{i,t} dV_{i,t}$. In particular, the stochastic integral $\int Z_t dW_t$ means $\sum_{l=1}^q \int Z_{l,t} dW_{l,t}$. When the integrator V is $\mathbb{R}^k$ valued and integrand U is $(\mathbb{R}^{n\times k})^\top$-valued, we write $\int U_t dV_t$ to mean the $\mathbb{R}^n$-valued random variable whose i-th component is $\int U_{i,t} dV_t$, where U_i is the i-th row of U.

For any vector $x \in \mathbb{R}^n$, $|x|$ is the vector 2-norm, defined by $(\sum_{i=1}^n |x_i|^2)^{1/2}$, and for any matrix A, $|A|$ is the matrix 2-norm, defined by $\max_{|x|=1} |Ax|$, where $|Ax|$ is the vector 2-norm of the vector Ax.

Let $\gamma \in (0,1]$ and $A(\cdot)$ be a function in the domain $[0,T) \times \mathbb{R}^l$ taking values in $\mathbb{R}^{k\times n}$ (resp. $\mathbb{R}^k$). We say that $A(t,\cdot)$ is γ-Hölder continuous uniformly in t with Hölder constant L_A if, for all $(x,y) \in (\mathbb{R}^l)^2$ and $t \in [0,T)$, $|A(t,x) - A(t,y)| \leq L_A|x-y|^\gamma$; in the case that $\gamma = 1$, we say that $A(t,\cdot)$ is Lipschitz continuous uniformly in t with Lipschitz constant L_A. Likewise, we say that $A(\cdot,x)$ is γ-Hölder continuous uniformly in x with Hölder constant L_A if, for every $(t_1,t_2) \in [0,T)^2$ and $x \in \mathbb{R}^l$, $|A(t_1,x) - A(t_2,x)| \leq L_A|t_1 - t_2|^\gamma$. We say that $A(t,\cdot)$ is continuously differentiable if the partial derivative $\partial_{x_j} A_{u,v}(t,x)$ (resp. $\partial_{x_j} A_u(t,x)$) of the (u,v)-th component $A_{u,v}(t,x)$ (reps. u-th component $A_u(t,x)$) of $A(t,x)$ exists and is continuous for every $(u,v) \in \{1,\dots,k\}\times\{1,\dots,n\}$, $j \in \{1,\dots,l\}$ and $(t,x) \in [0,T)\times\mathbb{R}^l$. If $A(t,\cdot)$ takes values in $\mathbb{R}^k$ and is differentiable, we define by $\nabla_x A(t,\cdot)$ the $\mathbb{R}^{k\times l}$ valued function whose (u,v)-th component is $\partial_{x_v} A_u(t,\cdot)$. If $A(t,\cdot)$ takes values in $(\mathbb{R}^k)^\top$ and is differentiable, we define by $\nabla_x A(t,\cdot)$ the $\mathbb{R}^{l\times k}$-valued function whose (u,v)-th component is $\partial_{x_u} A_v(t,\cdot)$. Define by $\|A\|_\infty$ the infinity norm

$$\max_{u,v} \sup_{(t,x)\in[0,T)\times\mathbb{R}^l} |A_{u,v}(t,x)| \quad (\text{resp.} \quad \max_u \sup_{(t,x)\in[0,T)\times\mathbb{R}^l} |A_u(t,x)|).$$

We will make use of a conditional version of Fubini's theorem, stated in Lemma 3.6.2. In order to simplify notation throughout this chapter, we write $\int_0^T \mathbb{E}_t[f_s]ds := \int_0^T F_t(\cdot,s)ds$, where F_t is the process defined in Lemma 1.8.1.

For a given time-grid $\pi = \{0 = t_0 < \ldots < t_N = T\}$, define the i-th time-increment by $\Delta_i := t_{i+1} - t_i$. For a parameter $\beta \in (0,1]$, the time grids $\pi_N^{(\beta)}$ are those with N time-points whose

i-th time point is $t_i = T - T(1 - i/N)^{1/\beta}$ for $i \in \{0, \dots, N\}$.

At several points, it will be necessary to apply a mollification procedure to continuous functions. The following definitions will come in handy.

Definition 1.1.1. *Let $M > 0$ and $R > 0$ be finite, and n a non-zero integer. A mollifier is a smooth function $\phi : \mathbb{R}^n \to [0, \infty)$ with compact support on $\{x :\in \mathbb{R}^n \ : \ |x| \leq 1\}$ such that $\int_{\mathbb{R}^n} \phi(x)dx = 1$ and $\lim_{M\to\infty}(RM)^n\phi(RMx) = \delta(x)$ for all $x \in \mathbb{R}^n$, where $\delta(x)$ is the Dirac delta function. Let $\phi_{R,M} : \mathbb{R}^n \to [0, \infty)$ be the function $x \mapsto (RM)^n\phi(RMx)$.*

An example of a mollifier is $\phi(x) = e^{-1/(1-|x|)}\mathbf{1}_{|x|<1}$. The following lemma demonstrates how a mollifier can be used to generate a smooth function from a continuous one, and shows that all continuous functions can be approximated by smooth functions.

Lemma 1.1.2. *Let $F : \mathbb{R}^n \to \mathbb{R}$ be continuous, and define the function $F_M(x) := \int_{\mathbb{R}^n} F(x - y)\phi_M(y)dy$. Then the function $F_M(x)$ is smooth and $\lim_{M\to\infty} F_M(x) = F(x)$ for all $x \in \mathbb{R}^n$.*

The proof is standard, but we include it for completeness.

Proof. Let $E_{R,M}$ be the compact support of $\phi_{R,M}$. Fix $x \in \mathbb{R}^n$.

$$|F(x) - F_M(x)| = |\int_{\mathbb{R}^n} (F(x) - F(x - y))\phi_{R,M}(y)dy| \leq \int_{E_{R,M}} |F(x) - F(x - y)|\phi_{R,M}(y)dy.$$

Since $\sup_{z,y\in E_{R,M}} |y - z|$ converges to zero as M increases to infinity, the result follows from the continuity of F.

1.1.2 Assumptions

The following assumptions will hold throughout this chapter.

($\mathbf{A}_{\mathbf{b},\sigma}$) The coefficients of the SDE (3.1.2) satisfy the following properties.

(i) $(t, x) \in [0, T]\times\mathbb{R}^d \mapsto b(t, x)$ is $\mathbb{R}^d$-valued, measurable and uniformly bounded. Moreover, $b(t, \cdot)$ is twice continuously differentiable with uniformly bounded derivatives and Hölder continuous second derivative, and $b(\cdot, x)$ is 1/2-Hölder continuous uniformly in x.

(ii) $(t, x) \in [0, T] \times \mathbb{R}^d \mapsto \sigma(t, x)$ is $\mathbb{R}^{d\times q}$-valued, measurable and uniformly bounded. Moreover, $\sigma(t, \cdot)$ is twice continuously differentiable with uniformly bounded derivatives and Hölder continuous second derivative, and $\sigma(\cdot, x)$ is 1/2-Hölder continuous uniformly in x.

($\mathbf{A}_{\mathrm{u.e.}}$) $\sigma(\cdot)$ satisfies a uniformly elliptic condition: there exists some finite $\bar{\beta} > 0$ such that, for any $\zeta \in \mathbb{R}^d$, $\zeta^\top\sigma(t, x)\sigma(t, x)^\top\zeta \geq \bar{\beta}|\zeta|^2$ for all $(t, x) \in [0, T] \times \mathbb{R}^d$.

($\mathbf{A}_{\mathbf{\Phi}}$) The terminal condition $\Phi : \mathbb{R}^d \to \mathbb{R}$ of the BSDE (y, z) in (3.1.1) is measurable and there is an $\alpha \in (0, 1]$ such that $\Phi(X_T) \in \mathbf{L}_{2,\alpha}$.

($\mathbf{A}_{\mathbf{f}}$) The driver $f : [0, T) \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^q \to \mathbb{R}$ satisfies (1.1.3).

In what follows, unless otherwise explicitly stated, we shall term a finite, non-negative real value depending only on x_0, the bounds of b and σ and their partial derivatives, $\bar{\beta}$, θ_L, θ_c, L_f, C_f, $K^\alpha(\Phi)$ and T a **constant** and denote it by C. If the time-grid $\pi^{(\beta)}$ is in use, a constant will also be allowed to depend on the parameter β.

The following assumptions will be needed for interim results only hold when specifically stated.

($\mathbf{A_{\partial f}}$) The driver $(t,x,y,z) \mapsto f(t,x,y,z)$ is continuously differentiable with respect (x,y,z) for all $t \in [0,T)$, and the partial derivatives are bounded by $L_f(T-t)^{(\theta_L-1)/2}$.

($\mathbf{A_{b\Phi}}$) The function Φ is uniformly bounded in the sense that $|\Phi(x)| \leq K$ for some $K \geq 0$.

($\mathbf{A_{exp\Phi}}$) The terminal condition has exponential bounds in the sense that there is a finite $C_\xi > 0$ such that $\mathbb{E}[e^{|\Phi(X_T)|}] \leq C_\xi$.

If ($\mathbf{A_{exp\Phi}}$) is in force, a constant C will also be allowed to depend on $\mathbb{E}[\exp(|\Phi(X_T)|)]$.

In the proofs below, it will be necessary to compute a right-inverse to the matrix $\sigma(\cdot)$, i.e., for every $(t,x) \in [0,T) \times \mathbb{R}^d$, it will be necessary to find a (q,d)-dimensional matrix $\sigma^{-1}(t,x)$ such that $\sigma(t,x)\sigma^{-1}(t,x) = I_d$. In the case where the dimensions d and q are equal, this is uniquely defined by usual matrix inverse of $\sigma(t,x)$, whose existence is guaranteed by the uniform ellipticity condition ($\mathbf{A}_{\text{u.e.}}$). If the dimensions d and q are not equal, $\sigma^{-1}(t,x)$ is defined by the pseudoinverse $\sigma(t,x)^\top\big(\sigma(t,x)\sigma(t,x)^\top\big)^{-1}$; this is well defined because the uniform ellipticity condition ($\mathbf{A}_{\text{u.e.}}$) guarantees the existence of the inverse of $\sigma\sigma^\top$. In both cases, the right-inverse is uniformly Lipschitz continuous. The following lemma is critical for showing this.

Lemma 1.1.3. *Let $\xi > 0$ be finite and $A : \mathbb{R}^n \to \mathbb{R}^{l\times l}$ be symmetric and such that $\eta^\top A(x)\eta \geq \xi|\eta|^2$ for all $x \in \mathbb{R}^n$ and $\eta \in \mathbb{R}^l$. Then, for every $x \in \mathbb{R}^d$, the matrix $A(x)$ is invertible and $|A^{-1}(x)| \leq 1/\xi$. Moreover, if $x \mapsto A(x)$ is γ-Hölder continuous,then it's inverse $x \mapsto A^{-1}(x)$ is also γ-Hölder continuous.*

Proof. Due to the condition $\eta^\top A(x)\eta \geq \xi|\eta|^2$, it follows that $A(x)$ is positive definite for every $x \in \mathbb{R}^n$. This implies that the singular values of $A(x)$ are all greater than ξ [GVL96, Theorem 8.1.2], and so $A(x)$ is invertible. Using the singular value decomposition of $A(x)$ to construct the inverse as in [GVL96, Section 5.5.4], the maximal sigular value of $A^{-1}(x)$ is less than $1/\xi$ and so, using [GVL96, Section 2.5.2] combined with the singular value decomposition of $A^{-1}(x)$, the matrix 2-norm of $A^{-1}(x)$ is equal to its maximal singular value, i.e. $|A^{-1}(x)| \leq 1/\xi$ for all $x \in \mathbb{R}^d$. Now, let x and y be elements in $\mathbb{R}^d$. Since $A^{-1}(y) - A^{-1}(x)$ is equal to

$$-A(x)^{-1}\big(A(y) - A(x)\big)A(y)^{-1},$$

it follows that

$$|A^{-1}(y) - A^{-1}(x)| \leq |A^{-1}(x)||A(y) - A(x)||A^{-1}(y)| \leq \frac{L_A}{\xi^2}|x-y|^\gamma,$$

where L_A is the Hölder constant of A. □

Lemma 1.1.4. *The right inverse matrix $\sigma(t,\cdot)^{-1}$ is Lipschitz continuous uniformly in t and $\sigma^{-1}(\cdot,x)$ is 1/2-Hölder continuous uniformly in x. Its Lipschitz (resp. Hölder) constant depends $\|\sigma\|_\infty$, $\|\nabla_x\sigma\|_\infty$ and $\bar\beta$ only, but not on (t,x). Moreover, $\|\sigma^{-1}\|_\infty \le \|\sigma\|_\infty/\bar\beta$.*

Proof. Let $t\in[0,T)$ be fixed, and define $A:[0,T)\times\mathbb{R}^d\to\mathbb{R}^{d\times d}$ by $A(t,x)=\sigma(t,x)\sigma(t,x)^\top$. It can be computed directly that $\sigma^{-1}(\cdot)=\sigma(\cdot)^\top A^{-1}(\cdot)$, whether or not d equals q. It follows from uniform ellipticity ($\mathbf{A}_{\text{u.e.}}$) and Lemma 1.1.3 that $|A^{-1}(t,x)|\le 1/\bar\beta$ for all $(t,x)\in[0,T)\times\mathbb{R}^d$. Due to the differentiability condition ($\mathbf{A}_{\mathbf{b},\sigma}$) on $\sigma(t,\cdot)$, $\sigma(t,\cdot)$ is Lipschitz continuous uniformly in t with Lipschitz constant $\|\nabla_x\sigma\|_\infty$, and, using additionally the equality $A(t,x)-A(t,y)=\sigma(y)(\sigma(t,x)^\top-\sigma(t,y)^\top)+(\sigma(t,x)-\sigma(t,y))\sigma(t,x)^\top$, $A(t,\cdot)$ is Lipschitz continuous uniformly in t with Lipschitz constant $2\|\sigma\|_\infty\|\nabla_x\sigma\|_\infty$. Using Lemma 1.1.3, it follows that $A^{-1}(t,\cdot)$ is Lipschitz continuous uniformly in t with Lipschitz constant $2\|\sigma\|_\infty\|\nabla_x\sigma\|_\infty/\bar\beta^2$. For any $(x,y)\in(\mathbb{R}^d)^2$, $\sigma(t,x)^{-1}-\sigma(t,y)^{-1}$ is equal to $\big(\sigma(t,x)^\top-\sigma(t,y)^\top\big)A^{-1}(t,x)+\sigma(t,y)^\top\big(A^{-1}(t,x)-A^{-1}(t,y)\big)$. and therefore

$$|\sigma(t,x)^{-1}-\sigma(t,y)^{-1}|\le\frac{\|\nabla_x\sigma\|_\infty}{\bar\beta}|x-y|+\frac{2\|\sigma\|_\infty\|\nabla_x\sigma\|_\infty}{\bar\beta^2}|x-y|.$$

The proof that $\sigma^{-1}(\cdot,x)$ is 1/2-Hölder continuous is essentially the same and we do not include it. $\square$

1.2 Malliavin calculus

We recall briefly some properties and definitions of Malliavin calculus which can be found in [Nua06, Chapter 1.2, Chapter 2.2].

Define $C_p^\infty(\mathbb{R}^m)$ to be the space of functions taking values in $\mathbb{R}$ which are infinitely differentiable such that all partial derivatives have at most polynomial growth, and denote by $W(h):=\int_0^T h_t dW_t$ the Itô integral of the $(\mathbb{R}^q)^\top$-valued, deterministic function $h\in\mathbf{L}_2([0,T];(\mathbb{R}^q)^\top)$. Let $\mathcal{R}\subset\mathbf{L}_2(\mathcal{F}_T)$ be the subspace containing all random variables of the form

$$F=f(W(h_1),\dots,W(h_m)),\quad f\in C_p^\infty(\mathbb{R}^m),\ h_i\in\mathbf{L}_2([0,T);\mathbb{R}^q)\quad\forall m\in\mathbb{N}.$$

Define the derivative operator $D:\mathcal{R}\mapsto\mathbf{L}_2([0,T]\times\Omega)$ by

$$D_tF:=\sum_{i=1}^m\partial_i f(W(h_1),\dots,W(h_m))h_i(t).$$

The derivative operator is extended to $\mathbb{D}^{1,2}\subset\mathbf{L}_2(\mathcal{F}_T)$, the closure of $\mathcal{R}$ in $\mathbf{L}_2(\mathcal{F}_T)$ under the norm

$$\|F\|_{1,2}:=\|F\|_2+\Big(\mathbb{E}[\int_0^T|D_tF|^2dt]\Big)^{1/2},$$

Define by $\mathbb{D}^{1,2}(\mathbb{R}^k)$ (resp. $\mathbb{D}^{1,2}((\mathbb{R}^k)^\top)$) by the space of random variables $F=(F_1,\dots,F_k)^\top$ (resp. $F=(F_1,\dots,F_k)$) such that $F_i\in\mathbb{D}^{1,2}$ for each $i\in\{0,\dots,k\}$. The Mallivin derivative DF is denoted by the $\mathbb{R}^{k\times q}$- (resp. $\mathbb{R}^{q\times k}$-) valued process whose i-th row (resp. column) is DF_i (resp. $(DF_i)^\top$).

The following lemma, termed the *chain rule* of Malliavin calculus, is proved in [Nua06, Proposition 1.2.3].

Lemma 1.2.1 (Chain rule). *Let $(F_1, \ldots, F_m) \in (\mathbb{D}^{1,2})^m$. For any continuously differentiable function $f : \mathbb{R}^m \to \mathbb{R}$ with bounded partial derivatives, and $F = f(F_1, \ldots, F_m) \in \mathbb{D}^{1,2}$, the random variable $f(F) \in \mathbb{D}^{1,2}$ and $Df(F) = \sum_{i=1}^m \partial_i f(F) DF_i = \nabla_x f(F) DF$.*

Remark. In the case that F takes values in $(\mathbb{R}^m)^\top$, the result of Lemma 1.2.1 hold with $Df(F) = \nabla_x f(F)(DF)^\top$. In the case that f takes values in $(\mathbb{R}^k)^\top$, applying Lemma 1.2.1 component-wise yields that $f(F)$ is in $\mathbb{D}^{1,2}((\mathbb{R}^k)^\top)$ and $Df(F) = (DF)^\top \nabla_x f(F)$.

For the space

$$\mathrm{dom}(\delta) := \{u \in \mathbf{L}_2([0,T] \times \Omega; (\mathbb{R}^q)^\top) : \exists c \in \mathbb{R} \; s.t \; \forall F \in \mathbb{D}^{1,2} \; |\mathbb{E}[\int_0^T (u_s \cdot D_s F) ds]| \leq c \|F\|_2^2 \}$$

define the Skorohod integral operator $\delta : \mathrm{dom}(\delta) \to \mathbf{L}_2(\Omega)$ as the dual operator to the Malliavin derivative in the sense that

$$\mathbb{E}[\int_0^T (u_s \cdot D_s F) ds] = \mathbb{E}[F\delta(u)].$$

The following Lemma, relating the Skorohod integral to Itô's integral, is given in [Nua06, Proposition 1.3.11]

Lemma 1.2.2 (Skorohod integral versus Itô integral). *If $u \in$ dom(δ) is adapted, $\delta(u) = \int_0^T u_s dW_s$, the Itô integral of u.*

The following Lemma, termed the *integration by parts formula* of Malliavin calculus, is proved in [Nua06, Chapter 1.3.1 (4)].

Lemma 1.2.3 (Integration by parts). *Suppose that $u \in$ dom(δ) and $F \in \mathbb{D}^{1,2}$ are such that $\mathbb{E}[F^2 \int_0^T |u_s|^2 ds] < \infty$. Then, the* integration by parts formula *holds:*

$$\int_0^T (u_s \cdot D_s F) ds = F\delta(u) - \delta(Fu).$$

This following result is used and proved in the proof of [MZ02, Theorem 4.2]. We include the proof here in greater detail for the convenience of the reader.

Lemma 1.2.4. *Let $u \in$ dom(δ). Then, for all $t \in [0,T)$, $\mathbb{E}_t[\delta(\mathbf{1}_{(t,T]}(\cdot)u)] = 0$.*

Proof. Let F be $\mathcal{F}_t$-measurable, square integrable random variable. We will show that $\mathbb{E}[F\delta(\mathbf{1}_{(t,T]}(\cdot)u.)] = 0$, whence the proof is completed by taking $F = \mathbf{1}_A$ for any $A \in \mathcal{F}_t$. Using [Nua06, Theorem 1.1.1], it suffices to prove that $\mathbb{E}[H_m\big(W(h)\big)\delta(\mathbf{1}_{(t,T]}(\cdot)u.)] = 0$ for the Hermite polynomial $H_m(\cdot)$ of degree m for any $m \geq 0$ and $h(\cdot)$ of the form $\tilde{h}(\cdot)\mathbf{1}_{[0,t]}(\cdot)$ for any bounded $\tilde{h} \in \mathbf{L}_2([0,T]; (\mathbb{R}^q)^\top)$. Suppose first that $m \geq 1$. The chain rule, Lemma 1.2.1 - yields $D_s H_m\big(W(h)\big) = H_{m-1}\big(W(h)\big)h(s)$ for every $s \in [0,T]$. Using the definition of the Skorohod integral,

$$\mathbb{E}[H_m\big(W(h)\big)\delta(\mathbf{1}_{(t,T]}(\cdot)u.)] = \mathbb{E}[\int_t^T u_s D_s H_m\big(W(h)\big) ds] = \mathbb{E}[\int_t^T u_s H_{m-1}\big(W(h)\big)h(s) ds] = 0$$

as required. The case $m = 0$ is simpler, because $H_0\big(W(h)\big)$ is 1 and so its Malliavin derivative is 0. $\square$

Remark 1.2.5. Suppose that the process u takes values in $\mathbb{R}^{q\times k}$ is such that $u_i^\top$ is in $\mathrm{dom}(\delta)$ for each $i \in \{0,\dots,k\}$, where u_i is the i-th column of u. The Skorohod integral of u, denoted by $\delta(u)$, is defined by

$$\delta(u) := \big(\delta(u_1^\top)^\top, \dots, \delta(u_k^\top)^\top\big). \tag{1.2.1}$$

In the case where u is adapted, Lemma 1.2.2 implies that $\delta(u)$ is equal to $(\int_0^T u_t^\top dW_t)^\top$. The integration by parts formula, Lemma 1.2.3, is applied column-wise in the case of matrix valued u. It follows that, if $F \in \mathbb{D}^{1,2}$ and u satisfy $\mathbb{E}[|F|^2 \int_0^T |u_t|^2 dt]$ is finite, then

$$\int_0^T (D_s F u_s) ds = F\delta(u) + \delta(Fu)$$

where $D_s F u_s$ is understood as a matrix-matrix multiplication, and the Skorohod integrals are defined in the multidimensional sense of equation (1.2.1).

Finally, we define the space

$$\begin{aligned}\mathbb{L}_1^2 := \Big\{ & u \in \mathbf{L}_2([0,T]\times\Omega) \ : u_t \in \mathbb{D}^{1,2} \ m-a.e., \\ & \exists \text{ measurable version of } (s,t,\omega) \mapsto D_s u_t(\omega) \ s.t. \\ & \mathbb{E}[\int_0^T \int_0^T |D_s u_t|^2 ds dt] < \infty \Big\}.\end{aligned}$$

1.2.1 SDEs and Malliavin calculus

We recall some standard properties on the Malliavin calculus applied to the SDE X defined in (1.1.2).

Lemma 1.2.6. *The random variables X_r - the marginals of the process X - are in $\mathbb{D}^{1,2}(\mathbb{R}^d)$ for each $r \in [0,T]$. For all s, $(D_s X_t)_{t\geq s}$ solves the linear SDE*

$$D_s X_t = \mathbf{1}_{[s,T]}(t)\Big\{\sigma(s,X_s) + \int_s^t \nabla_x b(r,X_r) D_s X_r dr + \sum_{j=1}^q \int_s^t \nabla_x \sigma_j(r,X_r) D_s X_r dW_{j,r}\Big\}.$$

where $\sigma_j(\cdot)$ is the j-th column of $\sigma(\cdot)$. Moreover, for every $p \geq 2$, there is a constant C_p depending only on $\|\nabla_x b\|_\infty$, $\|\sigma\|_\infty$, $\|\nabla_x \sigma_j\|_\infty$, T and p such that $\mathbb{E}[\sup_{s\leq r\leq T} |D_s X_r|^p] \leq C_p$.

Proof. The relation to the SDE is proved in [Nua06, Theorem 2.2.1]. The bound is proved using the proof method of [RY99, Theorem IX.2.4] (essentially using Gronwall's inequality). $\square$

The Malliavin derivative pocess $(D_s X_t)_{0\leq s\leq t\leq T}$ is related to the gradient process ∇X and its inverse $\nabla X^{(-1)}$, which are respectively defined by the SDEs

$$\nabla X_t = I_d + \int_0^t \nabla_x b(r, X_r)\nabla X_r dr + \sum_{j=1}^q \int_0^t \nabla_x \sigma_j(r, X_r)\nabla X_r dW_{j,r},$$

$$\nabla X_t^{(-1)} = I_d + \int_0^t \nabla X_r^{(-1)}(\nabla_x \sigma(r, X_r)^2 - \nabla_x b(r, X_r))dr - \sum_{j=1}^q \int_0^t \nabla X_r^{(-1)} \nabla_x \sigma_j(r, X_r) dW_{j,r},$$

where σ_j is the j-th column of σ. The notation and results of [Pro05, p325 - 327] are required to make the connection between the SDEs: for a $\mathbb{R}^{d\times d}$-valued continuous semimartingale $(\mathcal{Z}_t)_t$, recall that the *left* (resp. *right*) stochastic exponential $\mathcal{E}(\mathcal{Z})_t$ is the solution to the linear SDE

$$\mathcal{E}(\mathcal{Z})_t = I_d + \int_0^t \mathcal{E}(\mathcal{Z})_r d\mathcal{Z}_r \qquad \Big(\text{resp. } \mathcal{E}^R(\mathcal{Z})_t = I_d + (\int_0^t \mathcal{E}(\mathcal{Z})_r^\top d\mathcal{Z}_r^\top)^\top\Big).$$

Now, taking the $\mathbb{R}^{d\times d}$-valued, continuous semimartingales $(\mathcal{W}_t)_t$ given by

$$\mathcal{W}_t = \int_0^t \nabla_x b(r, X_r)dr + \sum_{j=1}^q \int_0^t \nabla_x \sigma_j(r, X_r) dW_{j,r}.$$

it follows that

$$\nabla X_t = \mathcal{E}^R(\mathcal{W})_t, \quad \nabla X_t^{(-1)} = \mathcal{E}(-\mathcal{W} + <\mathcal{W}>)_t, \tag{1.2.2}$$

where $<\mathcal{W}>$ is the quadratic variation process of $\mathcal{W}$.

Lemma 1.2.7. *For every $p \geq 2$, ∇X and $\nabla X^{(-1)}$ are in $\mathcal{S}^p$, and there is a constant C_p depending only on $\|\nabla_x b\|_\infty$, $\|\nabla_x \sigma_j\|_\infty$, T and p such that $\|\nabla X\|_{\mathcal{S}^p} + \|\nabla X^{(-1)}\|_{\mathcal{S}^p} \leq C_p$. Moreover, $\|\nabla X_t^{(-1)} - \nabla X_s^{(-1)}\|_2^2 \leq C|t-s|$ for $(t,s) \in [0,T]^2$. The processes $(\nabla X_t)_t$ and $(\nabla X_t^{(-1)})_t$ satisfy $\nabla X_t \nabla X_t^{(-1)} = I_d$ for all t a.s. Moreover, for all $0 \leq s, t \leq T$,*

$$D_s X_t = \nabla X_t \nabla X_s^{(-1)} \sigma(s, X_s)\mathbf{1}_{[s,T]}(t) \quad a.s.$$

Proof. That ∇X and $\nabla X^{(-1)}$ are in $\mathcal{S}^p$, and the bounds, can be proved using the proof method of [RY99, Theorem IX.2.4] (essentially using Gronwall's inequality). That $\nabla X_t \nabla X_t^{(-1)}$ is the identity for all t almost surely follows immediately from [Pro05, Chapter 5 Theorem 48]. To prove the second property, observe from (1.2.2) that $\nabla X_t = \nabla X_s + (\int_s^t (\nabla X_r)^\top \mathcal{W}_r^\top)^\top$. Using the invertibility property, it follows that $(\nabla X_t \nabla X_s^{(-1)} \sigma(s, X_s))_{t\geq s}$ solves the SDE

$$\mathcal{X}_t = \sigma(s, X_s) + (\int_s^t (\mathcal{X}_r)^\top \mathcal{W}_r^\top)^\top.$$

If one expands out the definition of $\mathcal{W}_r$, one sees that the process $\big(\nabla X_t \nabla X_s^{(-1)} \sigma(s, X_s)\mathbf{1}_{[s,T]}(t)\big)_{t\geq s}$ solves the same linear SDE as $(D_s X_t)_{t\geq s}$ given in Lemma 1.2.6, whence the result follows from the uniqueness of solutions of linear SDEs. □

A direct consequence of Lemma 1.2.7 is the following result.

Lemma 1.2.8. *For any $p \geq 2$, it holds that* $\sup_s \mathbb{E}[\sup_{s\leq r\leq T} |D_s X_r|^2]^{1/2} \leq C$.

1.2.2 FBSDEs and Malliavin calculus

Definition 1.2.9. *Let $\varepsilon \in [0,T)$ and define $f^{(\varepsilon)}(t,x,y,z) := f(t,x,y,z)\mathbf{1}_{[0,T-\varepsilon)}(t)$. Let $(Y^{(\varepsilon)}, Z^{(\varepsilon)})$ be the solution of the BSDE*

$$Y_t^{(\varepsilon)} = \Phi(X_T) + \int_t^T f^{(\varepsilon)}(s, X_s, Y_s^{(\varepsilon)}, Z_s^{(\varepsilon)})ds - \int_t^T Z_s^{(\varepsilon)} dW_s. \tag{1.2.3}$$

Additionally, let (y,z) be the solution of the BSDE with zero driver $y_t = \Phi(X_T) - \int_t^T z_s dW_s$ and $(y^{(\varepsilon)}, z^{(\varepsilon)})$ the solution of the BSDE with zero terminal condition

$$y_t^{(\varepsilon)} = \int_t^T f^{(\varepsilon)}(s, X_s, y_s + y_s^{(\varepsilon)}, z_s + z_s^{(\varepsilon)})ds - \int_t^T z_s^{(\varepsilon)} dW_s. \tag{1.2.4}$$

Since $f^{(\varepsilon)}(t,x,y,z)$ is Lipschitz continuous uniformly in t with Lipschitz constant $L_f\varepsilon^{(\theta_L-1)/2}$, the solutions of the BSDEs in Definition 1.2.9 exists in $\mathcal{S}^2 \times \mathcal{H}^2$ and are unique for all $\varepsilon \in [0,T)$ [EKPQ97, Theorem 2.1]. The introduction of the BSDEs in Definition 1.2.9 is localization technique used extensively in [GM10], and we shall frequently take advantage of it throughout this work. We shall also make use of the decomposition $(Y^{(\varepsilon)}, Z^{(\varepsilon)}) = (y + y^{(\varepsilon)}, z + z^{(\varepsilon)})$, which is standard in BSDE literature [GM10].

We now briefly outline some Malliavin calculus and path properties of the BSDEs (y,z) and $(y^{(\varepsilon)}, z^{(\varepsilon)})$. The results are proven in [GM10] and shall be used extensively throughout this work.

We first treat the linear BSDE (y,z). This BSDE is strongly related to the linear PDE

$$\left.\begin{array}{l} 0 = \partial_t u + \frac{1}{2}\sum_{i,j=1}^d (\sigma\sigma^\top)_{i,j} \frac{\partial^2}{\partial_{x_i}\partial_{x_j}} u + \sum_{i=1}^d b_i \frac{\partial}{\partial_{x_i}} u, \\ u(T,x) = \Phi(x). \end{array}\right\} \tag{1.2.5}$$

The following Lemma relates the linear BSDE (y,z) to the PDE in (1.2.5) and gives some boundedness properties for the function u and its derivatives.

Lemma 1.2.10. *Let $(\mathbf{A_{b\Phi}})$ be in force. Then $u(t,x) = \mathbb{E}[\Phi(X_T)|X_t = x]$ is a classical solution of the PDE (1.2.5). u, $\nabla_x u$, $\nabla_x^2 u$, $\nabla_x^3 u$, $\partial_t u$, $\partial_t \nabla_x u$ exist and are continuous. There is a constant C such that $|\nabla_x u(t,x)| \le C\|\Phi\|_\infty (T-t)^{-1/2}$ and $|\nabla_x^2 u(t,x)| \le C\|\Phi\|_\infty (T-t)^{-1}$ for all $(t,x) \in [0,T) \times \mathbb{R}^d$. Moreover, $\big(u(t,X_t), (\nabla_x u(t,X_t)\sigma(t,X_t))^\top\big)$ is the solution to the linear BSDE (y,z).*

Proof. The existence of u and its partial derivatives is standard [GM10, Introduction, part (c)]. To prove the boundedness of the $\nabla_x u$ and $\nabla_x^2 u$, observe from the proof of [GM10, Lemma 1.1] that

$$\nabla_x u(t, X_t) = \mathbb{E}_t[g(X_T)H_{t,T}^{(1)}], \quad \nabla_x^2 u(t,X_t) = E_t[\Phi(X_T)H_{t,T}^{(2)}]$$

where $H_{t,T}^{(1)}$ and $H_{t,T}^{(2)}$ are random variables satisfying

$$\mathbb{E}_t[|H_{t,T}^{(1)}|^2] \le C(T-t)^{-1}, \qquad \mathbb{E}_t[|H_{t,T}^{(2)}|^2] \le C(T-t)^{-2}.$$

The proof is now completed using the Cauchy-Schwartz inequality combined with the boundedness of Φ. $\square$

The following Lemma is a direct consequence of the existence and boundedness of the partial derivatives of u, given in Lemma 1.2.10, and the chain rule, Lemma 1.2.1.

Lemma 1.2.11. *Let* $(\mathbf{A_{b\Phi}})$ *be in force. Then, for all* $s \leq t$,

$$D_s y_t = \nabla_x u(t, X_t) D_s X_t \text{ and } D_s z_t = (D_s X_t)^\top U(t, X_t), \tag{1.2.6}$$

$$\text{for } U(t,x) := \nabla_x^2 u(t,x)\sigma(t,x) + \sum_{j=1}^{d} (\nabla_x u)_j(t,x) \nabla_x \sigma_j^\top(t,x), \tag{1.2.7}$$

where $(\nabla_x u)_j$ *is the* j*-th component of* $\nabla_x u$, *and* $\sigma_j^\top$ *is the* j*-th row of* σ. *Moreover,* $z_t = D_t y_t$ *for all* t, *and*

$$D_s y_t = \nabla_x u(t, X_t) \nabla X_t \nabla X_s^{(-1)} \sigma(s, X_s) \text{ and } D_s z_t = \big(\nabla X_t \nabla X_s^{(-1)} \sigma(s, X_s)\big)^\top U(t, X_t). \tag{1.2.8}$$

Proof. The representation $z_t = D_t y_t$ holds because Lemma 1.2.6 yields that $D_t X_t = \sigma(t, X_t)$. Finally, (1.2.8) follows from (1.2.6) and Lemma 1.2.7. $\square$

We move onto the non-linear BSDE $(y^{(\varepsilon)}, z^{(\varepsilon)})$.

Lemma 1.2.12. *Let* $(\mathbf{A_{\partial f}})$ *and* $(\mathbf{A_{b\Phi}})$ *be in force. Then* $(y_t^{(\varepsilon)}, z_t^{(\varepsilon)}) \in \mathbb{L}_1^2 \times \mathbb{L}_1^2$, *and that, for every* $0 \leq s < T$, $(D_s y_t^{(\varepsilon)}, D_s z_t^{(\varepsilon)})_{s \leq t \leq T}$ *is a version of the solution of the linear BSDE*

$$\begin{aligned} Q_t^{(s)} = &\int_t^T f_x^{(\varepsilon)}(\Theta_r) D_s X_r + f_y^{(\varepsilon)}(\Theta_r)(\nabla_x u(r, X_r) D_s X_r + Q_r^{(s)}) dr \\ &+ \int_t^T f_z^{(\varepsilon)}(\Theta_r) U(r, X_r)^\top D_s X_r + \sum_{j=1}^{q} f_{j,z}^{(\varepsilon)}(\Theta_r)(P_{j,r}^{(s)})^\top dr - \sum_{j=1}^{q} \int_t^T (P_{j,r}^{(s)})^\top dW_{j,r} \end{aligned} \tag{1.2.9}$$

where $P_j^{(s)}$ *is the* j*-th column of* $P^{(s)}$, $f_{j,z}^{(\varepsilon)}(\cdot)$ *is the* j*-th component of* $f_z^{(\varepsilon)}(\cdot)$, $\Theta_r := (r, X_r, Y_r^{(\varepsilon)}, Z_r^{(\varepsilon)})$, *and the function* $U(\cdot)$ *defined in equation* (1.2.7) *of Lemma 1.2.11. Moreover,* $z_t^{(\varepsilon)} = D_t y_t^{(\varepsilon)} \quad m \times \mathbb{P} - a.e.$

Proof. The proof is analogous to [GM10, Lemma 2.2]. It is only necessary to update the bounds on the partial derivatives of $f^{(\varepsilon)}(t, \cdot)$ from $(\mathbf{A_{\partial f}})$ and the bounds on the partial derivatives of u from Lemma 1.2.10. $\square$

There is a strong relationship between $(D_s y_t^{(\varepsilon)}, D_s z_t^{(\varepsilon)})_{t \geq s}$ and the solution of the linear BSDE

$$\begin{aligned} \nabla y_t^{(\varepsilon)} = &\int_t^T f_x^{(\varepsilon)}(\Theta_r) \nabla X_r + f_y^{(\varepsilon)}(\Theta_r)(u(r, X_r) \nabla X_r + \nabla y_r^{(\varepsilon)}) dr \\ &+ \int_t^T f_z^{(\varepsilon)}(\Theta_r) U(r, X_r)^\top \nabla X_r + \sum_{j=1}^{q} f_{j,z}^{(\varepsilon)}(\Theta_r)(\nabla z_{j,r}^{(\varepsilon)})^\top dr - \sum_{j=1}^{q} \int_t^T (\nabla z_{j,r}^{(\varepsilon)})^\top dW_r. \end{aligned} \tag{1.2.10}$$

where $\nabla z_j^{(\varepsilon)}$ is the j-th column of $\nabla z^{(\varepsilon)}$, $f_{j,z}^{(\varepsilon)}(\cdot)$ is the j-th component of $f_z^{(\varepsilon)}(\cdot)$, and the function

$U(\cdot)$ defined in equation (1.2.6) of Lemma 1.2.11.

Lemma 1.2.13. *Let* $(\mathbf{A_{\partial f}})$ *and* $(\mathbf{A_{b\Phi}})$ *be in force. Then, for all* $s \in [0,T)$,

$$Q_t^{(s)}\sigma^{-1}(s,X_s)\nabla X_s = \nabla y^{(\varepsilon)}\mathbf{1}_{[s,T)}(\cdot) \quad \text{and} \quad \big(\sigma^{-1}(s,X_s)\nabla X_s\big)^{\top}P^{(s)} = \nabla z^{(\varepsilon)}\mathbf{1}_{[s,T)}(\cdot)$$

in $\mathbf{L}_2([0,T)\times\Omega)$, *whence*

$$(D_s y^{(\varepsilon)}\sigma^{-1}(s,X_s)\nabla X_s, \big(\sigma^{-1}(s,X_s)\nabla X_s\big)^{\top}D_s z^{(\varepsilon)}) = (\nabla y^{(\varepsilon)}\mathbf{1}_{[s,T)}(\cdot), \nabla z^{(\varepsilon)}\mathbf{1}_{[s,T)}(\cdot)) \quad m\times\mathbb{P}-a.e. \tag{1.2.11}$$

Proof. Using the representation $D_sX_r = \nabla X_r \nabla X_s^{(-1)}\sigma(r,X_r)\mathbf{1}_{[s,T]}(r)$ of Lemma 1.2.7, whence it follows that $D_sX_r\sigma^{-1}(s,X_s)\nabla X_s = \nabla X_r\mathbf{1}_{[s,T]}(r)$. Substitute this into (1.2.9) to determine (1.2.11).

□

The following representations and a priori estimates will be useful.

Lemma 1.2.14. *Let* $(\mathbf{A_{\partial f}})$ *and* $(\mathbf{A_{b\Phi}})$ *hold. Define* $\Theta_r = (r, X_r, Y_r^{(\varepsilon)}, Z_r^{(\varepsilon)})$, *and set*

$$\begin{aligned} a_r^{(\varepsilon)} &:= \nabla_x f^{(\varepsilon)}(\Theta_r) + \nabla_y f^{(\varepsilon)}(\Theta_r)\nabla_x u(r,X_r) + \nabla_z f(\Theta_r)U(r,x)^{\top}, \\ b_r^{(\varepsilon)} &:= \nabla_y f^{(\varepsilon)}(r,X_r,Y_r^{(\varepsilon)},Z_r^{(\varepsilon)}), \quad c_r^{(\varepsilon)} := \nabla_z f^{(\varepsilon)}(r,X_r,Y_r^{(\varepsilon)},Z_r^{(\varepsilon)}). \end{aligned} \tag{1.2.12}$$

where $U(r,x)$ *is defined in* (1.2.7) *in Lemma 1.2.11. Then there a constant* C *such that*

$$\|a_r^{(\varepsilon)}\|_2 \le C\mathbf{1}_{[0,T-\varepsilon)}(r)(T-r)^{(\alpha+\theta_L-3)/2} \tag{1.2.13}$$

There exists a unique solution $(U^{(\varepsilon)}, V^{(\varepsilon)}) \in \mathcal{S}^2\times\mathcal{H}^2$ *of the BSDE*

$$\begin{aligned} U_t^{(\varepsilon)} = &\int_t^T a_r^{(\varepsilon)} + U_r^{(\varepsilon)}(b_r^{(\varepsilon)}I_d + \nabla_x b(r,X_r) + \sum_{j=1}^q c_{j,r}^{(\varepsilon)}\nabla_x\sigma_j(r,X_r))dr \\ &+ \int_t^T \sum_{j=1}^q (V_{j,r}^{(\varepsilon)})^{\top}(c_{j,r}^{(\varepsilon)}I_d + \nabla_x\sigma_j(r,X_r))dr - \sum_{j=1}^q\int_t^T (V_{j,r}^{(\varepsilon)})^{\top}dW_{j,r} \end{aligned} \tag{1.2.14}$$

where $\sigma_j(\cdot)$ *is the* j*-th column of* $\sigma(\cdot)$, $c_{r,j}^{(\varepsilon)}$ *are the* j*-th component of* $c_r^{(\varepsilon)}$, *and* $V_{j,r}^{(\varepsilon)}$ *is the* j*-th column of* $V_r^{(\varepsilon)}$. *There is a (possibly different) constant* C *such that, for any* $0 \le t < T$ *and* $\varepsilon > 0$,

$$\mathbb{E}[\sup_{t\le r<T}|U_r^{(\varepsilon)}|^2] + \int_t^T \|V_r^{(\varepsilon)}\|_2^2 dr \le C\|\int_t^{T-\varepsilon}|a_r^{(\varepsilon)}|dr\|_2^2 \le \frac{C}{\varepsilon^{1-(\theta_L+\alpha)\wedge 1}}. \tag{1.2.15}$$

Moreover, $z^{(\varepsilon)}$ *and* $\nabla z^{(\varepsilon)}$ *satisfy*

$$z_t^{(\varepsilon)} = U_t^{(\varepsilon)}\sigma(t,X_t) \quad m\times\mathbb{P}-a.e. \tag{1.2.16}$$

$$(V_{j,t}^{(\varepsilon)})^{\top} = (\nabla z_{j,t}^{(\varepsilon)})^{\top}\sigma^{-1}(t,X_t) - U_t^{(\varepsilon)}\nabla_x\sigma_j(t,X_t)\; m\times\mathbb{P}-a.e. \tag{1.2.17}$$

where $\nabla z_{j,t}^{(\varepsilon)}$ *is the* j*-th column of* $\nabla z_t^{(\varepsilon)}$.

Proof. From [GM10, Lemma 1.1], $\|\nabla_x u(t,X_t)\|_2 \leq C(T-t)^{(\alpha-1)/2}$ and $\|\nabla_x^2 u(t,X_t)\|_2 \leq C(T-t)^{(\alpha-2)/2}$. Therefore,

$$(\int_0^{T-\varepsilon} \|a_r^{(\varepsilon)}\|_2 dr)^2 < \frac{C}{\varepsilon^{1-(\theta_L+\alpha)\wedge 1}} < \infty.$$

This is the second inequality in (1.2.15). Additionally, for all $t \in [0,T)$, $|b_t^{(\varepsilon)}| + \max_j |c_{j,t}^{(\varepsilon)}| \leq C(T-t)^{(\theta_L-1)/2}$ almost surely. The first inequality in (1.2.15) and the existence and uniqueness of the solution now follow from Lemma 1.8.3.

The proofs of (1.2.16) and (1.2.17) are given in [GM10, Theorem 2.1]. The inclussion of the local Lipschitz continuity assumptions (1.1.3) make no difference, because the driver $f(t,x,y,z)\mathbf{1}_{[0,T-\varepsilon)}(t)$ is Lipschitz continuous uniformly in t in (x,y,z) with Lipschitz coefficient $L_f\varepsilon^{(\theta_L-1)/2}$. $\square$

1.3 Applications

1.3.1 Quadratic BSDE

Assume that $q = d$. Consider a quadratic growth driver satisfying

$$|f(t,x,y,z)| \leq L\ (1+|y|+|z|^2),$$
$$|f(t,x,y,z) - f(t,x',y',z')| \leq L\ (1+|z|+|z'|)(|x-x'|+|y-y'|+|z-z'|)$$

for any $(t,x,x',y,y',z,z') \in [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R} \times (\mathbb{R}^d)^\top \times (\mathbb{R}^d)^\top$ and for a given finite $L \geq 0$. Assume additionally that the terminal function Φ is Hölder continuous and bounded. Then [DG06, Theorem 2.1] yields that the continuous-time BSDE is given by $Y_t = u(t,X_t)$ and $Z_t = \nabla_x u(t,X_t)\sigma(t,X_t)$ where u is the unique solution to the semi-linear PDE $\partial_t u(t,x) + \mathcal{L}u(t,x) + f(t,x,u(t,x),\nabla_x u(t,x)\sigma(t,x)) = 0$ with $u(T,x) = \Phi(x)$. Moreover, there exist constants $\theta \in (0,1]$ and $C_u \in \mathbb{R}^+$ such that

$$(T-t)^{(1-\theta)/2}|\nabla u(t,x)\sigma(t,x)| \leq C_u, \quad \forall (t,x) \in [0,T) \times \mathbb{R}^d.$$

Now, set $\varphi_t : \zeta \in \mathbb{R} \mapsto \varphi_t(\zeta) = \operatorname{sign}(\zeta)\min\big(|\zeta|, \frac{C_u}{(T-t)^{(1-\theta)/2}}\big)$ and define the new driver $\bar{f}(t,x,y,z) := f\big(t,x,y,\varphi_t(z_1),\ldots,\varphi_t(z_d)\big)$. Observe that $\bar{f}(t,X_t,Y_t,Z_t) = f(t,X_t,Y_t,Z_t)$, thus the BSDE with driver f and $\bar{f}$ have the same solution. Notice also that $\varphi_t(\cdot)$ is 1-Lipschitz continuous and bounded by $\frac{C_u}{(T-t)^{(1-\theta)/2}}$, hence $\bar{f}(t_i,X_{t_i},y,z)$ satisfies (1.1.3) with $C_f = L$, $\theta_c = 1$, $L_f = L(T^{(1-\theta)/2} + 2\sqrt{d}C_u)$, $\theta_L = \theta$. One may argue that the construction of $\bar{f}$ depends on the knowledge of C_u and θ. In Remark 1.4.7 of Section 1.4, we demonstrate that θ is in fact equal to the Hölder exponent α, whereas C_u can be expressed in terms of C_f, L_f, T, θ_c, θ_L, and the bound and Hölder constant of Φ.

1.3.2 BSDE decomposition and proxy methods

Consider the case $\theta_L = 1$, and assume we explicitly know $(v(t,x),\nabla v(t,x)\sigma(t,x))$, the solution to the linear parabolic equation $\partial_t v(t,x) + \tilde{\mathcal{L}}v(t,x) + \tilde{f}(t,x) = 0$; the diffusion process associated to $\tilde{\mathcal{L}}$, the terminal condition and the driver may have changed to produce an analytical solution. v

is called *proxy* in [BGM09]. It is then natural to numerically compute the residual $(Y_t^0, Z_t^0) := (Y_t - v(t, X_t), Z_t - \nabla v(t, X_t)\sigma(t, X_t))$. It solves a BSDE with terminal function $\Phi(.) - v(T, .)$ and driver

$$f^0(t, x, y, z) := f(t, x, y + v(t, x), z + \nabla v(t, x)\sigma(t, x)) - \tilde{f}(t, x) + (\mathcal{L} - \tilde{\mathcal{L}})v(t, x).$$

The new driver f^0 is uniformly Lipschitz w.r.t. y and z. If $v(T, .)$ is θ_Φ-Hölder continuous ($\theta_\Phi \in (0, 1]$), then usual PDE estimates on the parabolic operator $\tilde{\mathcal{L}}$ give $\sup_{t<T}(T-t)^{(\frac{k-\theta_\Phi}{2})_+}|\nabla_x^{(k)} v(t, x)| \leq C_v$ ($k = 0, 1, 2$), from which (1.1.3) is satisfied for f^0 with $\theta_c = \frac{1+\theta_\Phi}{2}$.

To complete this example, we mention that in the case $\tilde{\mathcal{L}} = \mathcal{L}$, $v(T, .) = \Phi(.)$ and $\tilde{f} = 0$. This has been investigated numerically in, for example, [BS12], where they include the solution $v(t, \cdot)$ as part of the numerical approximation of (Y^0, Z^0) and demonstrate that this may be more efficient than using a naïve method.

In general, even when no proxy is available, it is still useful to use the decomposition $v(T, \cdot) = \Phi(\cdot)$, $\tilde{\mathcal{L}} = \mathcal{L}$ of the BSDE and to approximate $v(t, \cdot)$ and $\nabla v(t, \cdot)$ at $t < T$. The numerical schemes for (Y^0, Z^0) are much better behaved due to zero terminal condition; this phenomenon is investigated in Chapter 3. Moreover, approximation of $(v(\cdot), \nabla v(\cdot))$ can be performed using a multilevel technique; see Chapter 3. This technique incorporates variance reduction and can be performed using parallel computing, leading to high efficiency of the overall method.

1.4 A priori estimates

Define the Malliavin weights

$$H_r^s = \frac{1}{r - s}\Big(\int_s^r (\sigma^{-1}(t, X_t)D_s X_t)^\top dW_t\Big)^\top, \quad 0 \leq s < r \leq T \tag{1.4.1}$$

where $D_s X_t$ is the Malliavin derivative of X_t at s defined in Section 1.2.1. It was shown in Lemma 1.1.4 that $|\sigma^{-1}(t, x)|$ is uniformly bounded in (t, x). The following constant appears throughout this chapter

$$C_M := \|\sigma^{-1}\|_\infty^2 \|\sigma\|_\infty^2 \|\nabla X\|_{\mathcal{S}^4}^2 \|\nabla X^{(-1)}\|_{\mathcal{S}^4}^2. \tag{1.4.2}$$

The following result is used in the proof of [GM10, Lemma 1.1]; we include it here for completeness.

Lemma 1.4.1. *For any* $0 \leq s \leq r \leq T$,

$$\mathbb{E}_s[|H_r^s|^2] \leq \frac{C_M}{r - s}$$

Moreover, for every $p \geq 2$, *there is a finite* $C_p \geq 0$ *depending only on* p, $\|\nabla_x b\|_\infty$, $\max_j \|\nabla_x \sigma_j\|_\infty$, *and* T *such that* $\|H_r^s\|_p \leq C_p(r - s)^{-p/2}$.

Proof. Observe, using Lemma 1.2.7 and the fact that $(s - r)^2|H_r^s|^2 - \int_s^r |\sigma^{-1}(t, X_t)D_s X_t|^2 dt$ is a (local) martingale, that

$$\mathbb{E}_s[|H_r^s|^2] = (r - s)^{-2}\mathbb{E}_s[\int_s^r |\sigma^{-1}(t, X_t)D_s X_t|^2 dt] \leq \frac{\|\sigma\|_\infty \|\sigma^{-1}\|_\infty}{(r - s)^2}\mathbb{E}_s[\int_s^r |\nabla X_t \nabla X_s^{(-1)}|^2 dt].$$

Moreover, since $\nabla X_t \nabla X_s^{(-1)} = I_d + (\int_s^t \left(\nabla X_u \nabla X_s^{(-1)}\right)^\top \mathcal{W}_u^\top)^\top$, as shown in the proof of Lemma 1.2.7, and since the partial derivatives of b and σ are uniformly bounded, it can easily be shown that $\mathbb{E}_s[|\nabla X_t \nabla X_s^{(-1)}|^2]$ is less than or equal to $\|\nabla X\|_{\mathcal{S}^4}^2 \|\nabla X^{(-1)}\|_{\mathcal{S}^4}^2$, whence the result follows.

The bound on $\|H_r^s\|_p$ is proved as above, using the Burkholder-Davis-Gundy inequality. $\square$

We now state and prove a priori results on the solutions of BSDEs with drivers satisfying (1.1.3).

Proposition 1.4.2. *Let $\Phi_1, \Phi_2 \in \mathbf{L}_2(\mathcal{F}_T)$ and $(\omega, t, y, z) \mapsto f_1(\omega, t, y, z), f_2(\omega, t, y, z)$ be $\mathcal{P} \otimes \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}((\mathbb{R}^q)^\top)$-measurable functions for which there are constants $(\theta_{1,L}, \theta_{2,L}) \in (0, 1]^2$ and $(L_{f_1}, L_{f_2}) \in (0, \infty)^2$ such that*

$$|f_i(\omega, t, y, z) - f_i(\omega, t, y', z')| \leq \frac{L_{f_i}\{|y - y'| + |z - z'|\}}{(T - t)^{(1-\theta_{i,L})/2}} \quad m \times \mathbb{P} - \textit{almost everywhere},$$

and $f_i(\omega, t, 0, 0) \in \mathcal{H}^2$ for $i \in \{1, 2\}$. Let (Y_i, Z_i) be a solution to the FBSDE with terminal condition Φ_i and driver $f_i(t, y, z)$ ($i = 1, 2$ respectively).

Define

$$\begin{aligned} &\Delta Y_t := Y_{1,t} - Y_{2,t}, \qquad \Delta Z_t := Z_{1,t} - Z_{2,t}, \\ &\Delta f_t := f_1(t, Y_{1,t}, Z_{1,t}) - f_2(t, Y_{1,t}, Z_{1,t}), \quad \Delta\Phi := \Phi_1 - \Phi_2. \end{aligned}$$

Then there is a constant C depending only on T, L_{f_2}, and $\theta_{2,L}$ such that

$$\sup_{0 \leq t \leq T} \|\Delta Y_t\|_2^2 + \int_0^T \|\Delta Z_t\|_2^2 dt \leq C\|\Delta\Phi\|_2^2 + C\| \int_0^T \Big|\Delta f_t\Big| dt\|_2^2 \tag{1.4.3}$$

Moreover, suppose that $Z_{i,t} := \mathbb{E}_t[\Phi_i H_T^t + \int_t^T f_i(r, Y_{i,r}, Z_{i,r}) H_r^t dr]$ for all $t \in [0, T)$ almost surely. Then there is a (possibly different) finite constant $C \geq 0$ depending only on T, C_M, L_{f_2}, and $\theta_{2,L}$ such that

$$\|\Delta Z_t\|_2 \leq C\frac{V_{t,T}(\Delta\Phi)}{\sqrt{T - t}} + C\int_t^T \frac{\|\Delta f_r\|_2}{\sqrt{r - t}} dr + C\|\Delta\Phi\|_2 (T - t)^{\theta_L/2} \quad \textit{for all } t \in [0, T), \tag{1.4.4}$$

where $V_{t,T}(\Delta\Phi) := \|\Delta\Phi - \mathbb{E}_i[\Delta\Phi]\|_2$.

Proof. In what follows, C depends only on T, L_{f_2}, C_M and $\theta_{2,L}$, and its value may change from line to line.

Using the definition of the BSDE (1.1.1),

$$\Delta Y_t + \int_t^T \Delta Z_s dW_s = \Delta\Phi + \int_t^T \Delta f_s ds + \int_t^T f_2(s, Y_{1,s}, Z_{1,s}) - f_2(s, Y_{2,s}, Z_{2,s}) ds.$$

Using (1.1.3) and Hölder's inequality,

$$\begin{aligned}
&\|\Delta Y_t\|_2^2 + \int_t^T \|\Delta Z_s\|_2^2 ds \\
&\leq 3\|\Delta\Phi\|_2^2 + 3\|\int_t^T \Delta f_s ds\|_2^2 \\
&+ 3\|\int_t^T f_2(s, Y_{1,s}, Z_{1,s}) - f_2(s, Y_{2,s}, Z_{2,s}) ds\|_2^2 \\
&\leq 3\|\Delta\Phi\|_2^2 + 3\|\int_t^T |\Delta f_s| ds\|_2^2 + 3L_{f_2}^2 \|\int_t^T \frac{|\Delta Y_s| + |\Delta Z_s|}{(T-s)^{(1-\theta_{2,L})/2}} ds\|_2^2 \\
&\leq 3\|\Delta\Phi\|_2^2 + 3\|\int_t^T |\Delta f_s| ds\|_2^2 \\
&+ 3L_{f_2}^2 \int_t^T \frac{1}{(T-s)^{1-\theta_{2,L}}} ds \int_t^T \{\|\Delta Y_s\|_2^2 + \|\Delta Z_s\|_2^2\} ds \\
&\leq 3\|\Delta\Phi\|_2^2 + 3\|\int_t^T |\Delta f_s| ds\|_2^2 + 3L_{f_2}^2 (T-t)^{\theta_{2,L}} \int_t^T \{\|\Delta Y_s\|_2^2 + \|\Delta Z_s\|_2^2\} ds
\end{aligned} \tag{1.4.5}$$

Setting $t_0 = (T - 1/(6L_{f_2}^2)^{1/\theta_{2,L}}) \vee 0$ ensures that $3L_{f_2}(T-t_0)^{\theta_{2,L}} \leq 1/2$, and, on the other hand, that $T - t_0 \leq 1$. Integrating (1.4.5) over (t_0, T), we obtain

$$\int_{t_0}^T \|\Delta Y_t\|_2^2 + \|\Delta Z_s\|_2^2 ds \leq 6\|\Delta\Phi\|_2^2 + 6L_{f_2}^2 \|\int_{t_0}^T |\Delta f_s| ds\|_2^2 \tag{1.4.6}$$

Substituting (1.4.6) into (1.4.5) then yields

$$\sup_{t_0 \leq t < T} \|\Delta Y_t\|_2^2 \leq 6\|\Delta\Phi\|_2^2 + 6\|\int_{t_0}^T |\Delta f_s| ds\|_2^2$$

and this gives the result in the interval $[t_0, T]$.

In the interval $[0, t_0)$, the function $(y, z) \mapsto f_2(\omega, t, y, z)$ is $m \times \mathbb{P}$ Lipschitz continuous with Lipschitz constant $L = L_f(T-t_0)^{(\theta_{2,L}-1)/2}$. It then follows from [EKPQ97, Proposition 2.1] that

$$\sup_{0 \leq t < t_0} \|\Delta Y_t\|_2^2 + \int_{t_0}^T \|\Delta Z_s\|_2^2 ds \leq C\|\Delta Y_{t_0}\|_2^2 + C\|\int_0^{t_0} |\Delta f_s| ds\|_2^2$$

and the proof of (1.4.3) is complete by substituting the bounds on $\|\Delta Y_{t_0}\|_2^2$ from above.

Next, we prove (1.4.4). Using the representation $Z_{i,t} = \mathbb{E}_t[\Phi_i H_T^t + \int_t^T f_i(r, Y_{i,r}, Z_{i,r}) H_r^t dr]$, it follows from Minkowski's inequality, the Cauchy-Schwarz inequality, and Lemma 1.4.1 that

$$\|\Delta Z_t\|_2 \leq \frac{CV_{t,T}(\Delta\Phi)}{\sqrt{T-t}} + C\int_t^T \frac{\|\Delta f_r\|_2}{\sqrt{r-t}} dr + C\int_t^T \frac{\|\Delta Y_r\|_2 + \|\Delta Z_r\|_2}{(T-r)^{(1-\theta_{2,L})/2}\sqrt{r-t}} dr. \tag{1.4.7}$$

Defining $\Theta_r = \|\Delta Y_r\|_2 + \|\Delta Z_r\|_2$ and recalling (1.4.5), it follows that

$$\Theta_t \leq C\|\Delta\Phi\|_2 + \frac{CV_{t,T}(\Delta\Phi)}{\sqrt{T-t}} + C\int_t^T \frac{\|\Delta f_r\|_2}{\sqrt{r-t}} dr + C\int_t^T \frac{\Theta_r}{(T-r)^{(1-\theta_{2,L})/2}\sqrt{r-t}} dr. \tag{1.4.8}$$

Applying Lemma 1.8.6 with $u_t = \Theta_t$ and

$$w_t = C\|\Delta\Phi\|_2 + \frac{CV_{t,T}(\Delta\Phi)}{\sqrt{T-t}} + C\int_t^T \frac{\|\Delta f_r\|_2}{\sqrt{r-t}}dr,$$

it follows that

$$\Theta_r \leq Cw_t + C\int_t^T \frac{w_r}{(T-r)^{(1-\theta_{2,L})/2}\sqrt{r-t}}dr + C\int_t^T \frac{\Theta_r}{(T-r)^{(1-\theta_{2,L})/2}}dr$$

whence it follows from Lemma 1.8.7 that

$$\int_t^T \frac{\Theta_r}{(T-r)^{(1-\theta_{2,L})/2}\sqrt{r-t}}dr \leq C\int_t^T \frac{w_r}{(T-r)^{(1-\theta_{2,L})/2}\sqrt{r-t}}dr$$

Substituting this into (1.4.7) and applying Lemma 1.8.5 leads to

$$\begin{aligned}
\|\Delta Z_t\|_2 &\leq \frac{CV_{t,T}(\Delta\Phi)}{\sqrt{T-t}} + C\int_t^T \frac{\|\Delta f_r\|_2}{\sqrt{r-t}}dr + C\int_t^T \frac{w_r}{(T-r)^{(1-\theta_{2,L})/2}\sqrt{r-t}}dr \\
&= \frac{CV_{t,T}(\Delta\Phi)}{\sqrt{T-t}} + C\int_t^T \frac{\|\Delta f_r\|_2}{\sqrt{r-t}}dr + C\int_t^T \frac{V_{r,T}(\Delta\Phi)}{(T-r)^{(2-\theta_{2,L})/2}\sqrt{r-t}}dr \\
&\quad + C\int_t^T \frac{\int_r^T \|\Delta f_s\|_2 (s-r)^{-1/2}ds}{(T-r)^{(1-\theta_{2,L})/2}\sqrt{r-t}}dr + C\|\Delta\Phi\|_2 (T-t)^{\theta_{2,L}/2} \\
&= \frac{CV_{t,T}(\Delta\Phi)}{\sqrt{T-t}} + C\int_t^T \frac{\|\Delta f_r\|_2}{\sqrt{r-t}}dr + C\int_t^T \frac{V_{r,T}(\Delta\Phi)}{(T-r)^{(2-\theta_{2,L})/2}\sqrt{r-t}}dr \\
&\quad + C\int_t^T \|\Delta f_s\|_2 \{\int_r^s (s-r)^{-1+\theta_{2,L}}(r-t)^{-1/2}dr\}ds + C\|\Delta\Phi\|_2 (T-t)^{\theta_{2,L}/2}.
\end{aligned}$$

The proof is completed by observing that $V_{r,T}(\Delta\Phi)$ is non-increasing. □

One can adapt the proof of [EKPQ97, Theorem 2.1] to prove the existence and uniqueness of the locally Lipschitz BSDE.

Corollary 1.4.3. *There exists a unique pair of process* (Y,Z) *in* $\mathcal{S}^2 \times \mathcal{H}^2$ *solving the BSDE* (1.1.1) *with terminal condition* $\Phi(X_T) \in \mathbf{L}_2(\mathcal{F}_T)$ *and driver* f *satisfiying the locally Lipschitz continuous and boundedness of* (1.1.3).

Proof. Let (ϕ,ψ) be in $\mathcal{H}^2 \times \mathcal{H}^2$, and define the random function

$$f(r,y,z) = f(r) := f(r, X_r, \phi_r, \psi_r).$$

The function f is predictably measurable and satisfies assumptions (H1)-(H5) of [BDH+03, Section 4]. Since f takes no argument in (y,z), it is only necessary to check (H1): using Minkowski's inequality, the Cauchy-Schwarz inequality, and the local Lipschitz continuity and boundedness (1.1.3), it follows that

$$\|\int_0^T |f(r)|dr\|_2 \leq \int_0^T \|f(r,X_r,0,0)\|_2 dr + L_f(\mathbb{E}[\int_0^T \{|\phi_r|^2 + |\psi_r|^2\}dr])^{1/2}(\int_0^T \frac{dr}{(T-r)^{1-\theta_L}})^{1/2} < \infty.$$

Thanks to [BDH+03, Theorem 4.2], there exists a unique solution $(Y^{(\phi,\psi)}, Z^{(\phi,\psi)})$ to the BSDE

$$Y_t^{(\phi,\psi)} = \Phi(X_T) + \int_t^T f(r)dr - \sum_{j=1}^q \int_t^T Z_{j,r}^{(\phi,\psi)} dW_{j,r}.$$

in $\mathcal{H}^2 \times \mathcal{H}^2$. The function $\Xi : \mathcal{H}^2 \times \mathcal{H}^2 \to \mathcal{H}^2 \times \mathcal{H}^2$ mapping (ϕ,ψ) to $(Y^{(\phi,\psi)}, Z^{(\phi,\psi)})$ is well defined. In fact, $Y^{(\phi,\psi)}$ is in $\mathcal{S}^2$. As in the proof of [EKPQ97, Theorem 2.1], we prove that Ξ is a contraction.

For $k \in \{1,2\}$, let $(\phi_k,\psi_k) \in \mathcal{H}^2 \times \mathcal{H}^2$ and define the BSDE $(Y_k, Z_k) := \Xi(\phi_k,\psi_k)$. Define the differences $\delta Y = Y_1 - Y_2$, $\delta Z = Z_1 - Z_2$, $\delta\phi = \phi_1 - \phi_2$ and $\delta\psi = \psi_1 - \psi_2$. It then follows exactly as in (1.4.5) that

$$\begin{aligned}\|\delta Y_t\|_2^2 + \int_t^T \|\delta Z_r\|_2 dr &\le \| \int_t^T |f(r, X_r, \phi_{1,r}, \psi_{1,r}) - f(r, X_r, \phi_{2,r}, \psi_{2,r})| dr\|_2^2 \\ &\le L_f^2 (T-t)^{\theta_L} \int_t^T \{\|\delta\phi_r\|_2^2 + \|\delta\psi_r\|_2^2\} dr\end{aligned}$$

for all $t \in [0,T)$. Setting $t_0 = (T - 1/(4L_f^2)^{1/\theta_L} \wedge 1) \vee 0$ ensures, on the one hand, that $L_f^2(T-t_0)^{\theta_L} \le 1/4$, and, on the other hand, that $T - t_0 \le 1$. Integrating the above inequality on the interval $t \in [t_0, T)$ then yields

$$\begin{aligned}\int_{t_0}^T \{\|\delta Y_r\|_2^2 + \|\delta Z_r\|_2^2\} dr &\le \frac{1}{4} \int_{t_0}^T \{\|\delta\phi_r\|_2^2 + \|\delta\psi_r\|_2^2\} dr, \\ \|\delta Y_t\|_2 &\le \frac{1}{4} \int_{t_0}^T \{\|\delta\phi_r\|_2^2 + \|\delta\psi_r\|_2^2\} dr \quad \text{for all } t \in [t_0, T).\end{aligned}$$

On the interval $[0, t_0)$, the function $f(t,x,\cdot)$ is Lipschitz continuous with a uniform Lipschitz constant for all (t,x), so we proceed as in the proof of Theorem [EKPQ97, Theorem 2.1] to show that, for sufficiently large $\eta > 0$,

$$\int_0^{t_0} e^{\eta r} \{\|\delta Y_r\|_2^2 + \|\delta Z_r\|_2^2\} dr \le e^{\eta t_0} \|\delta Y_{t_0}\|_2 + \frac{1}{2} \int_0^{t_0} e^{\eta r} \{\|\delta\phi_r\|_2^2 + \|\delta\psi_r\|_2^2\} dr$$

Combining this with the above estimates on $\int_{t_0}^T \{\|\delta Y_r\|_2^2 + \|\delta Z_r\|_2^2\} dr$ and $\|\delta Y_{t_0}\|_2$ then yields

$$\int_0^T e^{\eta_r} \{\|\delta Y_r\|_2^2 + \|\delta Z_r\|_2^2\} dr \le \frac{1}{2} \int_0^T e^{\eta_r} \{\|\delta\phi_r\|_2^2 + \|\delta\psi_r\|_2^2\} dr$$

where $\eta_r = \eta(r \wedge t_0)$. This is sufficient to prove that Ξ is a contraction. □

The a priori estimates (1.4.4) allow us to determine bounds on the second moments of Z_t.

Corollary 1.4.4 (Moment bounds)**.** *Assume that f satisfies* (1.1.3)*, and $Z_t = \mathbb{E}_t[\Phi(X_T) H_T^t + \int_t^T f(r, X_r, Y_r, Z_r) H_r^t dr]$ for all $t \in [0,T)$ almost surely. Then there is a constant C such that, for* *all* $t \in [0,T)$,

$$\sup_{0\le t\le T}\|Y_t\|_2 \le C, \quad \textit{and} \quad \|Z_t\|_2 \le \frac{C}{(T-t)^{(1-\alpha)/2}} + \frac{C}{(T-t)^{(1-2\theta_c)/2}} + C(T-t)^{\theta_L/2}. \tag{1.4.9}$$

Moreover, for all $t \in [0,T)$,

$$\|f(s,X_s,Y_s,Z_s)\|_2 \le \frac{C}{(T-s)^{1-((2\theta_c)\wedge\alpha+\theta_L)/2}} + \frac{C}{(T-s)^{1-\theta_c}}. \tag{1.4.10}$$

Proof. In what follows, C is a constant that depends only on L_f, C_M, θ_L, θ_c, C_f, T and $K^\alpha(\Phi)$ whose value may change from line to line.

Apply (1.4.4) from Proposition 1.4.2 with $(Y_1,Z_1)=(0,0)$ and $(Y_2,Z_2)=(Y,Z)$ to obtain (for all $t\in[0,T)$)

$$\begin{aligned}\|Z_t\|_2 &\le C\frac{V_{t,T}(\Phi)}{\sqrt{T-t}} + C\int_t^T \frac{\|f(r,X_r,0,0)\|_2}{\sqrt{r-t}}dr + C\|\Phi\|_2(T-t)^{\theta_L/2}\\ &\le \frac{C}{(T-t)^{(1-\alpha)/2}} + C\int_t^T \frac{dr}{(T-r)^{1-\theta_c}\sqrt{r-t}} + C(T-t)^{\theta_L/2}\end{aligned}$$

and the (1.4.9) follows. Combining the local Lipschitz continuity and boundedness of f in (1.1.3) and the bounds in (1.4.9) leads to (1.4.10). □

It is also possible to generalize the a priori estimates (1.4.4) of Proposition 1.4.2 using conditional expectations; for simplicity, the following proposition is stated in the setting that $f_i(t,y,z)$ is $f_i(t,X_t,y,z)$ and Φ_i is $\Phi_i(X_T)$ for $i\in\{1,2\}$.

Lemma 1.4.5. *Let* $x\mapsto\Phi_1(x),\Phi_2(x)$ *be measurable functions such that* $\big(\Phi_1(X_T),\Phi_2(X_T)\big)\in\big(\mathbf{L}_2(\mathcal{F}_T)\big)^2$*. Additionally, let* $(t,x,y,z)\mapsto f_1(t,x,y,z), f_2(t,x,y,z)$ *be functions satisfying*

$$|f_i(t,x,y,z)-f_i(t,x',y',z')| \le L_{f_i}\frac{|x-x'|+|y-y'|+|z-z'|}{(T-t)^{(1-\theta_{i,L})/2}}, \quad |f_i(t,0,0,0)| \le \frac{C_{f_i}}{(T-t)^{1-\theta_{i,C}}}$$

with parameters $(L_{f_i},C_{f_i})\in(0,\infty)^2$ *and* $(\theta_{i,L},\theta_{i,C})\in(0,1]^2$ *(*$i=1,2$ *respectively), and denote by* (Y_i,Z_i) *the solution to the FBSDE with terminal condition* $\Phi_i(X_T)$ *and driver* $f_i(t,x,y,z)$ *(*$i=1,2$ *respectively). Then there is a finite constant* $C\ge0$ *depending only on* T*,* L_{f_2} *and* $\theta_{2,L}$ *such that, for all* $s<t$*,*

$$\mathbb{E}_s[\Delta Y_t^2] \le C\mathbb{E}_s[\Delta\Phi^2] + C\Big(\int_t^T \mathbb{E}_s[\Delta f_r^2]^{1/2}dr\Big)^2 \tag{1.4.11}$$

Moreover, suppose that $Z_{i,t} := \mathbb{E}_t[\Phi_i(X_T)H_T^t + \int_t^T f_i(r,X_r,Y_{i,r},Z_{i,r})H_r^t dr]$ *for all* $t\in[0,T)$ *almost surely (*$i=1,2$*). Then there is a (possibly different) finite constant* $C\ge0$ *depending only on* T*,* C_M*,* L_{f_2}*, and* $\theta_{2,L}$ *such that,*

$$\mathbb{E}_s[|\Delta Z_t|^2]^{1/2} \le C\frac{\mathbb{E}_s\big[(\Delta\Phi-\mathbb{E}_t\Delta\Phi)^2\big]^{1/2}}{\sqrt{T-t}} + C\int_t^T \frac{\mathbb{E}_s[\Delta f_r^2]^{1/2}}{\sqrt{r-t}}dr + C\mathbb{E}_s[\Delta\Phi^2]^{1/2}(T-t)^{\theta_L/2} \tag{1.4.12}$$

for all $t \in [0,T)$ *almost surely.*

The proof of Lemma 1.4.5 is the analogous to the proof of Proposition 1.4.2; the only difference is that one must use the conditional version of the Minkowski, Cauchy-Schwarz, and Hölder inequalities and the conditional Fubini's theorem (Lemma 1.8.1). Lemma 1.4.5 is particularly useful when the terminal condition is bounded and Hölder continuous, because this allows us to make almost sure estimates on $|Z_t|$.

Corollary 1.4.6. *Assume that* f *satisfies* (1.1.3)*, for some* $\alpha \in (0,1]$*,* Φ *is bounded and* α*-Hölder continuous, and* $Z_t = \mathbb{E}_t[\Phi(X_T)H_T^t + \int_t^T f(r, X_r, Y_r, Z_r)H_r^t dr]$ *for all* $t \in [0,T)$ *almost surely. Then there is a constant* C *such that*

$$|Z_t| \leq \frac{C\|\Phi\|_\infty}{(T-t)^{(1-\alpha)/2}} + \frac{C}{(T-t)^{(1-2\theta_c)/2}} + C\|\Phi\|_\infty (T-t)^{\theta_L/2} \quad \text{for all } t \in [0,T) \text{ almost surely.}$$

Proof. The proof is analogous to the proof of Corollary 1.4.4, but we use Lemma 1.4.5 in the place of Proposition 1.4.2. □

Remark 1.4.7. We can use Corollary 1.4.6 to specify the coefficient θ_L of the quadratic BSDE in Section 1.3.1. In particular, since $\theta_c = 1$ in this case, we see that $|Z_t| \leq C(T-t)^{(\alpha-1)/2}$, so it suffices to set $\theta_L = \alpha$.

Recall $(Y^{(\varepsilon)}, Z^{(\varepsilon)})$ from Definition 1.2.9 in Section 1.2.2, the BSDE with terminal condition $\Phi(X_T)$ and driver $f^{(\varepsilon)}(t,x,y,z) := f(t,x,y,z)\mathbf{1}_{[0,T-\varepsilon)}(t)$. The following corollary of Proposition 1.4.2 will be used extensively throughout this chapter; it provides a stability results between the BSDEs (Y,Z) and $(Y^{(\varepsilon)}, Z^{(\varepsilon)})$ that are controlled by ε.

Corollary 1.4.8. *Let* $\gamma := (\theta_c \wedge \frac{\alpha}{2} + \frac{\theta_L}{2}) \wedge \theta_c$*. Suppose that* $Z = \mathbb{E}_t[\Phi(X_T)H_T^t + \int_t^T f(s, X_s, Y_s, Z_s)H_s^t ds]$ *and* $Z_t^{(\varepsilon)} = \mathbb{E}_t[\Phi(X_T)H_T^t + \int_t^T f^{(\varepsilon)}(s, X_s, Y_s^{(\varepsilon)}, Z_s^{(\varepsilon)})H_s^t ds]$ *for all* $t \in [0,T)$ *almost surely. Then there is a constant* C *such that*

$$\sup_{0\leq t\leq T} \|Y_t - Y_t^{(\varepsilon)}\|_2^2 + \int_0^T \|Z_t - Z_t^{(\varepsilon)}\|_2^2 dt \leq C\varepsilon^{2\gamma}, \tag{1.4.13}$$

$$\|Z_t - Z_t^{(\varepsilon)}\|_2 \leq C \int_{t\vee(T-\varepsilon)}^T \frac{ds}{(T-s)^{1-\gamma}\sqrt{s-t}} \tag{1.4.14}$$

for all $t \in [0,T)$*. In particular,* $(Y^{(\varepsilon)}, Z^{(\varepsilon)}) \to (Y,Z)$ *as* $\varepsilon \to 0$ *in* $\mathcal{S}^2 \times \mathcal{H}^2$*.*

Proof. In what follows, C may change from line to line.

It follows from (1.4.3) in Proposition 1.4.2 that

$$\sup_{0\leq t\leq T} \|Y_t - Y_t^{(\varepsilon)}\|_2^2 + \int_0^T \|Z_s - Z_s^{(\varepsilon)}\|_2^2 ds \leq C\Big(\int_{T-\varepsilon}^T \|f(s, X_s, Y_s, Z_s)\|_2 ds\Big)^2. \tag{1.4.15}$$

Substituting (1.4.10) into (1.4.15) combined with $\Big(\int_{T-\varepsilon}^T \frac{ds}{(T-s)^{(1-\gamma)}}\Big)^2 \leq C\varepsilon^{2\gamma}$ completes the proof of (1.4.13). Next, it follows from (1.4.4) that

$$\|Z_t - Z_t^{(\varepsilon)}\|_2 \leq C\int_{t\vee(T-\varepsilon)}^T \frac{\|f(s, X_s, Y_s, Z_s)\|_2}{\sqrt{s-t}} ds \quad \text{for all } t \in [0,T).$$

Substituting (1.4.10) above proves (1.4.14). □

To end this section, we present a mollification procedure that will be used frequently to allow us to extend results under the assumptions ($\mathbf{A_{\partial f}}$) and ($\mathbf{A_{b\Phi}}$) to the same results without these assumptions.

Corollary 1.4.9. *Let $M > 0$ and $t \mapsto R_t > 1$ be finite for all $t \in [0,T)$. Define $\Phi_M(x) := -M \vee \Phi(x) \wedge \Phi(x)$ and, recalling the function $\phi_{R,M}$ from Definition 1.1.1,*

$$f_M(t,x,y,z) := \int_{\mathbb{R}^d \times \mathbb{R} \times (\mathbb{R}^q)^\top} f(t, x-x', y-y', z-z')\phi_{R_t,M}(x',y',z')d(x',y',z'),$$

and let (Y_M, Z_M) be the solution of the BSDE with terminal condition $\Phi_M(X_T)$ and driver $f_M(t, X_t, y, z)$. Then Φ_M satisfies ($\mathbf{A_{b\Phi}}$), f_M satisfies ($\mathbf{A_{\partial f}}$), and $(Y_M, Z_M) \to (Y, Z)$ as $M \to \infty$ in $\mathcal{S}^2 \times \mathcal{H}^2$.

Proof. Applying Proposition 1.4.2 with $(Y_1, Z_1) = (Y, Z)$ and $(Y_2, Z_2) = (Y_M, Z_M)$, it follows that there is a constant $C \geq 0$ such that

$$\|Y - Y_M\|^2_{\mathcal{S}^2} + \int_0^T \|Z_t - Z_{M,t}\|_2^2 dt \leq C\|\Phi(X_T) - \Phi_M(X_T)\|_2^2 + C(\int_0^T \|f(t,X_t,Y_t,Z_t) - f_M(t,X_t,Y_t,Z_t)\|_2 dt)^2.$$

for a constant C that does not depend on M. Since $\Phi_M(x) \to \Phi(x)$ and $f_M(t,x,y,z) \to f(t,x,y,z)$ as $M \to \infty$ for all $(t,x,y,z) \in [0,T) \times \mathbb{R}^d \times \mathbb{R} \times (\mathbb{R}^q)^\top$, it follows that $\Phi_M(X_T) \to \Phi(X_T)$ and $f_M(t,X_t,Y_t,Z_t) \to f(t,X_t,Y_t,Z_t)$ - Lemma 1.1.2 - for all t almost surely as $M \to \infty$. Therefore $(Y_M, Z_M) \to (Y, Z)$ as $M \to \infty$ in $\mathcal{S}^2 \times \mathcal{H}^2$ follows from the dominated convergence theorem.

Suppose that $F : \mathbb{R}^n \to \mathbb{R}$ is a Lipschitz continuous function with Lipschitz constant L. Then for any $(x, x') \in (\mathbb{R}^n)^2$, it follows from the definition of $\phi_{R,M}$ and the properties of convolutions that

$$|F_M(x) - F_M(x')| \leq \int_{\mathbb{R}^n} \phi_{R,M}(y)|f(x-y) - f(x'-y)|dy \leq L|x-y|$$

so $F_M(x)$ is also Lipschitz continuous with Lipschitz constant L. Since $F_M(\cdot)$ is smooth, this implies that the partial derivatives of F_M are absolutely bounded by L. Therefore, the partial derivatives of $f_M(t,x,y,z)$ are bounded by $L_f(T-t)^{(\theta_L - 1)/2}$. □

The following Lemma follows directly from Proposition 1.4.2.

Corollary 1.4.10. *Let $M > 0$ be finite, define the function $\Phi_M(x) := -M \vee \Phi(x) \wedge M$, and let $(\bar{Y}_M, \bar{Z}_M)$ be the solution to the BSDE with terminal condition $\Phi_M(X_T)$ and driver $f(t, X_t, y, z)$. Then Φ_M satisfies ($\mathbf{A_{b\Phi}}$), and there is a constant C such that*

$$\|Y - \bar{Y}_M\|_{\mathcal{S}^2} + \int_0^T \|Z_t - \bar{Z}_{M,t}\|_2^2 dt \leq C\|\Phi(X_T) - \Phi_M(X_T)\|_2^2.$$

If additionally $\bar{Z}_{M,t}$ is equal to $\mathbb{E}_t[\Phi_M(X_T)H_T^t + \int_t^T f(s, X_s, \bar{Y}_{M,s}, \bar{Z}_{M,s})H_s^t ds]$ for all $t \in [0,T)$ almost surely, there is a (possibly different) constant C such that

$$\|Z_t - \bar{Z}_{M,t}\|_2 \leq \frac{C\|\Phi(X_T) - \Phi_M(X_T)\|_2}{\sqrt{T-t}} \quad \textit{for all } t \in [0,T).$$

1.5 Representation theorem

Theorem 1.5.1. *Suppose that* $\Phi \in L_{2,\alpha}$ *and* $(t,x,y,z) \mapsto f(t,x,y,z)$ *satisfies* (1.1.3). *Then, there is a predictable version* $\mathcal{Z}$ *of* Z *which satisfies*

$$\mathcal{Z}_t = \mathbb{E}_t[\Phi(X_T)H_T^t + \int_t^T f(s,X_s,Y_s,Z_s)H_s^t ds] \quad \text{for all } t \in [0,T) \quad \mathbb{P}-a.s. \tag{1.5.1}$$

where H_s^t *are the Malliavin weights given in* (1.4.1).

Proof. In the following, we C is a constant whose value may change from line to line.

To start with, assume $(\mathbf{A_{\partial f}})$ and $(\mathbf{A_{b\Phi}})$, and recall $(Y^{(\varepsilon)}, Z^{(\varepsilon)})$, the BSDE (1.2.3) defined is Section 1.2.2. We will use the results of Section 1.2.2 - in particular, the representation of the Malliavin derivatives $(D.y^{(\varepsilon)}, D.y^{(\varepsilon)})$ of Lemma 1.2.13 and Lemma 1.2.11 - to determine the representation formula for $(Y^{(\varepsilon)}, Z^{(\varepsilon)})$.

To make use of (1.2.11), we make some computations on the Malliavin derivative of $f(\Theta_r)$. By the chain rule Lemma 1.2.1,

$$\begin{aligned} D_v f^{(\varepsilon)}(\Theta_r) &= f_x^{(\varepsilon)}(\Theta_r)D_vX_r + f_y^{(\varepsilon)}(\Theta_r)D_vY_r^{(\varepsilon)} + f_z^{(\varepsilon)}(\Theta_r)(D_vZ_r^{(\varepsilon)})^\top \\ &= f_x^{(\varepsilon)}(\Theta_r)D_vX_r + f_y^{(\varepsilon)}(\Theta_r)(\nabla_x u(r,X_r)D_vX_r + D_v y_r^{(\varepsilon)}) \\ &\quad + f_z^{(\varepsilon)}(\Theta_r)(U(r,X_r)^\top D_vX_r + (D_v z_r^{(\varepsilon)})^\top) \end{aligned}$$

for all $0 \le v \le r$. We use the representation (1.2.11) (with $s = v$) to show that for all $0 \le v \le r$,

$$\begin{aligned} D_v f^{(\varepsilon)}(\Theta_r)\sigma^{-1}(v,X_v)\nabla X_v &= f_x^{(\varepsilon)}(\Theta_r)\nabla X_r + f_y^{(\varepsilon)}(\Theta_r)(\nabla_x u(r,X_r)\nabla X_r + \nabla y_r^{(\varepsilon)}) \\ &\quad + f_z^{(\varepsilon)}(\Theta_r)(U(r,X_r)^\top \nabla X_r + (\nabla z_r^{(\varepsilon)})^\top) \qquad m \times \mathbb{P}-a.e. \end{aligned}$$

Since the expression on the right hand side is independent of v, we can integrate over $v \in (t,r)$ to obtain

$$\begin{aligned} \frac{1}{r-t}\int_t^r D_v f^{(\varepsilon)}(\Theta_r)\sigma^{-1}(v,X_v)\nabla X_v dv &= f_x^{(\varepsilon)}(\Theta_r)\nabla X_r + f_y^{(\varepsilon)}(\Theta_r)(\nabla_x u(r,X_r)\nabla X_r + \nabla y_r^{(\varepsilon)}) \\ &\quad + f_z^{(\varepsilon)}(\Theta_r)(U(r,X_r)^\top \nabla X_r + (\nabla z_r^{(\varepsilon)})^\top) \qquad m \times \mathbb{P}-a.e. \end{aligned} \tag{1.5.2}$$

We address the integral expression above by applying the integration by parts rule of Malliavin calculus (Lemma 1.2.3) with $u. = \sigma^{-1}(\cdot, X.)\nabla X.$ and $F = f^{(\varepsilon)}(\Theta_r)$. It has been shown in [MZ02, Theorem 4.2], $(\sigma^{-1}(\cdot, X.)\nabla X.) \in \mathrm{dom}(\delta)$. We still need to prove

$$\mathbb{E}[|f^{(\varepsilon)}(\Theta_r)|^2 \int_t^T |\sigma^{-1}(u,X_u)\nabla X_u|^2 du] < \infty \quad dr-a.e. \tag{1.5.3}$$

Since $\nabla X_s \in \mathcal{S}^p$ for all $p \ge 2$ and σ is uniformly elliptic, we have that $\int_t^T |\sigma^{-1}(u,X_u)\nabla X_u|^2 du \in \mathbf{L}_4(\mathcal{F}_T)$. We would like to apply the Cauchy-Schwarz inequality, but we cannot assume a priori that $f^{(\varepsilon)}(\Theta_r) \in \mathbf{L}_4(\mathcal{F}_T)$. We use (1.1.3) to obtain the decomposition

$$|f^{(\varepsilon)}(\Theta_r)| \leq \frac{C_f}{\varepsilon^{1-\theta_c}} + \frac{L_f}{\varepsilon^{(1-\theta_L)/2}}(|Y_r^{(\varepsilon)}| + |Z_r^{(\varepsilon)}|)$$

whence it is sufficient to show that $Y_r^{(\varepsilon)}, Z_r^{(\varepsilon)} \in \mathbf{L}_4(\mathcal{F}_T)$. Lemma 1.4.5 implies that

$$|Y_r^{(\varepsilon)}| \leq C\mathbb{E}_r[|\Phi(X_T)|^2]^{1/2} + C.$$

But Φ is bounded, so we have that $\mathbb{E}[|Y_r^{(\varepsilon)}|^4] \leq C\|\Phi\|_\infty < \infty$ as required. Lemma 1.2.11 and 1.2.12 give us that $Z_r^{(\varepsilon)} = D_r Y_r^{(\varepsilon)} = D_r y_r + D_r y_r^{(\varepsilon)} \quad m \times \mathbb{P} - a.e.$ We treat $D_r y_r$ and $D_r y_r^{(\varepsilon)}$ separately. Setting $s = r$ and applying [GM10, (A.2) in Lemma A.1] to the nonlinear BSDE (1.2.9),

$$\begin{aligned}
\mathbb{E}[|D_r y_r^{(\varepsilon)}|^4] &\leq C\mathbb{E}[\Big(\int_r^{T-\varepsilon} L_f |D_r X_v|(1 + |\nabla_x u(v, X_v)| + |\nabla_x^2 u(v, X_v)|)(T-v)^{(\theta_L - 1)/2} dv\Big)^4] \\
&\leq \frac{C\|\Phi\|_\infty \mathbb{E}[\sup_{r\leq v\leq T} |D_r X_v|^4]}{\varepsilon^8} < \infty
\end{aligned}$$

follows from the bounds of the derivatives of $f^{(\varepsilon)}$, the bounds on the derivatives of u in Lemma 1.2.10 and the bounds on $\mathbb{E}[\sup_{r\leq v\leq T} |D_r X_v|^4]$ from Lemma 1.2.8.

Using that $D_r y_r = \nabla_x u(r, X_r)\sigma(r, X_r)$ and the bounds of the derivatives of u from Lemma 1.2.10, $\mathbb{E}[|D_r y_r|^4] \leq C\|\sigma\|_\infty/\varepsilon^2 < \infty$. Hence, we have shown that $Z_r^{(\varepsilon)} \in \mathbf{L}_4(\mathcal{F}_T)$. This concludes the proof of (1.5.3). We can now apply integration by parts (Remark 1.2.5) to obtain

$$\begin{aligned}
&\int_t^r D_v f^{(\varepsilon)}(\Theta_r)\sigma^{-1}(v, X_v)\nabla X_v dv \\
&\quad = f^{(\varepsilon)}(\Theta_r)\delta(\mathbf{1}_{[t,r]}(\cdot)\sigma^{-1}(\cdot, X_\cdot)\nabla X_\cdot) \; - \; \delta\big(\mathbf{1}_{[t,r]}(\cdot) f^{(\varepsilon)}(\Theta_r)\sigma^{-1}(\cdot, X_\cdot)\nabla X_\cdot\big) \\
&\quad = f^{(\varepsilon)}(\Theta_r)(\int_t^r (\sigma^{-1}(v, X_v)\nabla X_v)^\top dW_v)^\top \; - \; \delta\big(\mathbf{1}_{[t,r]}(\cdot) f^{(\varepsilon)}(\Theta_r)\sigma^{-1}(\cdot, X_\cdot)\nabla X_\cdot\big) \qquad (1.5.4)
\end{aligned}$$

where the first Skorohod integral in the first equality equals the Itô integral because $(\sigma^{-1}(v, X_v)\nabla X_v)$ is adapted (Remark 1.2.5).

We now return to (1.2.10) and apply the conditional expectation $\mathbb{E}_t[\cdot]$ combined with the conditional Fubini's theorem, Lemma 1.8.1, to obtain

$$\begin{aligned}
\nabla y_t^{(\varepsilon)} &= \int_t^T \mathbb{E}_t[f_x^{(\varepsilon)}(\Theta_r)\nabla X_r + f_y^{(\varepsilon)}(\Theta_r)(u(r, X_r)\nabla X_r + \nabla y_r^{(\varepsilon)})]dr \\
&\quad + \int_t^T \mathbb{E}_t[f_z^{(\varepsilon)}(\Theta_r)(U(r, X_r)^\top \nabla X_r + (\nabla z_r^{(\varepsilon)})^\top)]dr
\end{aligned}$$

where the integrals on the right hand side have the meaning given in Lemma 1.8.1. Due to the $m \times \mathbb{P}$ equality of (1.5.2) and (1.5.4), we can write

$$\nabla y_t^{(\varepsilon)} = \int_t^T \mathbb{E}_t[f^{(\varepsilon)}(\Theta_r)\frac{\big(\int_t^r (\sigma^{-1}(v, X_v)\nabla X_v)^\top dW_v\big)^\top}{r-t}]dr + \int_t^T \frac{\mathbb{E}_t[\delta\big(\mathbf{1}_{[t,r]}(\cdot) f^{(\varepsilon)}(\Theta_r)\sigma^{-1}(\cdot, X_\cdot)\nabla X_\cdot\big)]}{r-t}dr$$

and, since the conditional expectation of the Skorkhod integral in the second integral is 0 for almost every r (Lemma 1.2.4), we obtain

$$\nabla y_t^{(\varepsilon)} = E_t[\int_t^T f^{(\varepsilon)}(\Theta_r)\frac{\left(\int_t^r(\sigma^{-1}(v,X_v)\nabla X_v)^\top dW_v\right)^\top}{r-t}dr].$$

We can now return to the representation (1.2.11) to see that $D_t y_t^{(\varepsilon)} = E_t[\int_t^T f^{(\varepsilon)}(\Theta_r)H_r^t dr]$ $m\times\mathbb{P}-a.e.$ Finally, since $z_t = \mathbb{E}_t[\Phi(X_T)H_T^t]$ $\quad m\times\mathbb{P}-a.e.$, we have that

$$Z_t^{(\varepsilon)} = \mathbb{E}_t[\Phi(X_T)H_T^t + \int_t^T f^{(\varepsilon)}(r,X_r,Y_r^{(\varepsilon)},Z_r^{(\varepsilon)})H_r^t dr] \quad m\times\mathbb{P}-a.e. \tag{1.5.5}$$

To complete the proof for $(Y^{(\varepsilon)},Z^{(\varepsilon)})$, define by $\mathcal{Z}^{(\varepsilon)}$ the predictable projection [JS03, Theorem 2.28] of the process $(\mathcal{X}_t^{(\varepsilon)} := \Phi(X_T)H_T^t + \int_t^T f^{(\varepsilon)}(r,X_r,Y_r^{(\varepsilon)},Z_r^{(\varepsilon)})H_r^t dr)_{t\in[0,T)}$, and observe from (1.5.5) that $Z_t^{(\varepsilon)} = \mathcal{Z}_t^{(\varepsilon)}$ $m\times\mathbb{P}$-almost everywhere.

Define by $\mathcal{Z}$ the predictable projection [JS03, Theorem 2.28] of the process $(\mathcal{X}_t := \Phi(X_T)H_T^t + \int_t^T f(r,X_r,Y_r,Z_r)H_r^t dr)_{\in[0,T)}$. We take the version of $Z^{(\varepsilon)}$ given by the the predictable representation of the process $(\mathcal{X}_t^{(\varepsilon)} := \Phi(X_T)H_T^t + \int_t^T f^{(\varepsilon)}(r,X_r,Y_r^{(\varepsilon)},Z_r^{(\varepsilon)})H_r^t dr)_{t\in[0,T)}$. We show first that $\|Z_t^{(\varepsilon)} - \mathcal{Z}_t\|_2 \to 0$ as $\varepsilon\to 0$ for almost all $t\in[0,T)$. Since $\|Z_t^{(\varepsilon)}\|_2^2$ is bounded above by $C(T-t)^{\alpha\wedge(2\theta_c)-1}$ for all t and ε - see the following paragraph for more details - this implies, by the dominated convergence theorem, that $Z^{(\varepsilon)}\to\mathcal{Z}$ in $\mathcal{H}^2$. On the other hand, $Z^{(\varepsilon)}\to Z$ in $\mathcal{H}^2$ was determined in Corollary 1.4.8, and therefore $Z_t = \mathcal{Z}_t$ $m\times\mathbb{P}-a.e.$, which completes the proof under the assumptions $(\mathbf{A_{\partial f}})$ and $(\mathbf{A_{b\Phi}})$.

We first need some intermediate bounds. Analogously to Corollary 1.4.4, the following bounds on the second moments of $Y^{(\varepsilon)}$ and $Z^{(\varepsilon)}$ hold:

$$\sup_{0\le r\le T}\|Y_r^{(\varepsilon)}\|_2^2 \le C, \quad \text{and} \quad \|Z_r^{(\varepsilon)}\|_2^2 \le \frac{C}{(T-r)^{1-\alpha\wedge(2\theta_c)}} \quad \text{for all } r\in[0,T).$$

We will use the notation $\gamma := (\frac{\alpha}{2}\wedge\theta_c + \frac{\theta_L}{2})\wedge\theta_c$ hereafter. It follows analogously to (1.4.10) that

$$\|f(r,X_r,Y_r^{(\varepsilon)},Z_r^{(\varepsilon)})\|_2 \le \frac{C}{(T-r)^{1-\gamma}} \quad \text{for all } r\in[0,T). \tag{1.5.6}$$

Fix $t\in[0,T)$ and $\eta>0$. Using the representation formula

$$Z_t^{(\varepsilon)} = \mathbb{E}_t[\Phi(X_T)H_T^t + \int_t^T f^{(\varepsilon)}(r,X_r,Y_r^{(\varepsilon)},Z_r^{(\varepsilon)})H_r^t dr],$$

it follows from Minkowski's inequality, the Cauchy-Schwarz inequality, and Lemma 1.4.1 that

$$\begin{aligned}\|Z_t^{(\varepsilon)} - \mathcal{Z}_t\|_2 &= \|\mathbb{E}_t[\int_t^T \big(f^{(\varepsilon)}(r,X_r,Y_r^{(\varepsilon)},Z_r^{(\varepsilon)}) - f(r,X_r,Y_r,Z_r)\big)H_r^t dr]\|_2 \\ &\le \|\mathbb{E}_t[\int_t^T \big(f^{(\varepsilon)}(r,X_r,Y_r^{(\varepsilon)},Z_r^{(\varepsilon)}) - f(r,X_r,Y_r^{(\varepsilon)},Z_r^{(\varepsilon)})\big)H_r^t dr]\|_2 \\ &\quad + C_M^{1/2}\int_t^T \frac{\|f(r,X_r,Y_r^{(\varepsilon)},Z_r^{(\varepsilon)}) - f(r,X_r,Y_r,Z_r)\|_2}{\sqrt{r-t}}dr. \end{aligned} \tag{1.5.7}$$

Taking $\varepsilon < (T-t)/2$ and using (1.5.6), it follows that

$$\|\mathbb{E}_t[\int_t^T \big(f^{(\varepsilon)}(r, X_r, Y_r^{(\varepsilon)}, Z_r^{(\varepsilon)}) - f(r, X_r Y_r^{(\varepsilon)}, Z_r^{(\varepsilon)})\big) H_r^t dr]\|_2 \leq C_M^{1/2} \int_{T-\varepsilon}^T \frac{\|f(r, X_r, Y_r^{(\varepsilon)}, Z_r^{(\varepsilon)})\|_2}{\sqrt{r-t}} dr$$
$$\leq \frac{C_M^{1/2}}{\sqrt{T-t-\varepsilon}} \int_{T-\varepsilon}^T \|f(r, X_r, Y_r^{(\varepsilon)}, Z_r^{(\varepsilon)})\|_2 dr \leq \frac{\sqrt{2C_M}}{\sqrt{T-t}} \int_{T-\varepsilon}^T \frac{dr}{(T-r)^{1-\gamma}} = \frac{\sqrt{2C_M}\varepsilon^\gamma}{\sqrt{T-t}}.$$

Taking $\varepsilon < \eta^{1/\gamma}(T-t)^{1/(2\gamma)}/(2C_M)^{1/(2\gamma)}$ is sufficient to bound the above term by η. On the other hand, let $\delta < (T-t)/2$

$$\begin{aligned} &\int_t^T \frac{\|f(r, X_r, Y_r^{(\varepsilon)}, Z_r^{(\varepsilon)}) - f(r, X_r, Y_r, Z_r)\|_2}{\sqrt{r-t}} dr \\ &\leq C_M^{1/2} \frac{\int_{t+\delta}^T \|f(r, X_r, Y_r^{(\varepsilon)}, Z_r^{(\varepsilon)}) - f(r, X_r, Y_r, Z_r)\|_2 dr}{\sqrt{\delta}} \\ &\quad + C_M^{1/2} \int_t^{t+\delta} \frac{\|f(r, X_r, Y_r^{(\varepsilon)}, Z_r^{(\varepsilon)}) - f(r, X_r, Y_r, Z_r)\|_2}{\sqrt{r-t}} dr \end{aligned} \tag{1.5.8}$$

To bound the first integral term in (1.5.8), we observe that

$$|f(r, X_r, Y_r^{(\varepsilon)}, Z_r^{(\varepsilon)}) - f(r, X_r, Y_r, Z_r)| \leq L_f\{|Y_r - Y_r^{(\varepsilon)}| + |Z_r - Z_r^{(\varepsilon)}|\}(T-r)^{(\theta_L - 1)/2}.$$

Applying Hölder's inequality,

$$\begin{aligned} &C_M^{1/2} \int_{t+\delta}^T \|f(r, X_r, Y_r^{(\varepsilon)}, Z_r^{(\varepsilon)}) - f(r, X_r, Y_r, Z_r)\|_2 dr \\ &\qquad \leq C_M^{1/2} L_f \Big(\int_0^T \frac{dr}{(T-r)^{1-\theta_L}}\Big)^{1/2} \Big(\sup_{0\leq s\leq T} \|Y_s - Y_s^{(\varepsilon)}\|_2^2 + \int_0^T \|Z_r - Z_r^{(\varepsilon)}\|_2^2 dr\Big)^{1/2} \end{aligned}$$

Using that $(Y^{(\varepsilon)}, Z^{(\varepsilon)}) \to (Y, Z)$ in $\mathcal{S} \times \mathcal{H}^2$ as $\varepsilon \to 0$ (Corollary 1.4.8), set ε sufficiently small so that the above is bounded above by $\sqrt{\delta}\eta$. To bound the second integral term in (1.5.8), use (1.4.10) and (1.5.6) to show that

$$\begin{aligned} &C_M^{1/2} \int_t^{t+\delta} \frac{\|f(r, X_r, Y_r^{(\varepsilon)}, Z_r^{(\varepsilon)}) - f(r, X_r, Y_r, Z_r)\|_2}{\sqrt{r-t}} dr \\ &\qquad \leq \frac{C_M^{1/2}}{(T-t-\delta)^{1-\gamma}} \int_t^{t+\delta} \frac{dr}{\sqrt{r-t}} \leq \frac{2^{1-\gamma} C}{(T-t)^{1-\gamma}} \sqrt{\delta} \end{aligned}$$

and set δ sufficiently small so that the above is bounded above by η. Therefore, we have shown that for almost every $t \in [0, T)$ and every $\eta > 0$, there is a sufficiently small ε such that $\|Z_t^{(\varepsilon)} - \mathcal{Z}_t\|_2 < 3\eta$. In other words, $\mathbb{E}[|Z_t^{(\varepsilon)} - \mathcal{Z}_t|^2] \to 0$ as $\varepsilon \to 0$ for every t, as required.

To prove the result without $(\mathbf{A_{\partial f}})$ and $(\mathbf{A_{b\Phi}})$, recall the mollified BSDE (Y_M, Z_M) from Corollary 1.4.9. Since Φ_M satisfies $(\mathbf{A_{b\Phi}})$ and f_M satisfies $(\mathbf{A_{\partial f}})$, there is a predictable version $\mathcal{Z}_M$ of Z_M satisfying $\mathcal{Z}_{M,t} = \mathbb{E}_t[\Phi_M(X_T)H_T^t + \int_t^T f_M(r, X_r, Y_{M,r}, Z_{M,r})H_r^t dr]$ for all $t \in [0, T)$ almost surely. Now, we can use analogous to the above arguments, together with the point-wise convergence of f_M to f and Φ_M to Φ, and the convergence of (Y_M, Z_M) to (Y, Z) in $\mathcal{S}^2 \times \mathcal{H}^2$ from

Corollary 1.4.9, to complete the proof in the general case. $\square$

1.6 Convergence rate of the $\mathbf{L}_2$-regularity

For a given time-grid $\pi := \{0 = t_0 < \ldots < t_N = T\}$, recall the definition of the $\mathbf{L}_2$-regularity given by

$$\mathcal{E}(\pi) := \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \|Z_t - \bar{Z}_{t_i}\|_2^2 dt, \qquad \bar{Z}_{t_i} := \frac{1}{\Delta_i}\mathbb{E}_t\big[\int_{t_i}^{t_{i+1}} Z_t dt\big]. \tag{1.6.1}$$

Following on from Section 1.5, we work with the version of Z given by (1.5.1) in Theorem 1.5.1, i.e.

$$Z_t = \mathbb{E}_t[\Phi(X_T)H_T^t + \int_t^T f(s, X_s, Y_s, Z_s)H_s^t ds] \quad \text{for all } t \in [0, T) \text{ almost surely.}$$

Since $(\bar{Z}_{t_i})_i$ is the projection of Z onto the space of adapted discrete processes with nodes on π under the scalar product $(u, v) = \mathbb{E}\int_0^T (u_s \cdot v_s)ds$, it follows that

$$\mathcal{E}(\pi) \leq \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \|Z_t - Z_{t_i}\|_2^2 dt.$$

In the case of uniformly Lipschitz drivers $\theta_L = 1$, [GM10] determine that FBSDEs with terminal conditions in $L_{2,\alpha}$ satisfy $\mathcal{E}(\pi^{(\beta)}) \leq CN^{-1}$ for the time-grids $\pi^{(\beta)} = \{0 = t_0 < \ldots < t_N = T\}$ given by $t_i := T - T(1 - i/N)^{1/\beta}$ and $\beta < \alpha$. We also define

$$\gamma := (\theta_c \wedge \frac{\alpha}{2} + \frac{\theta_L}{2}) \wedge \theta_c. \tag{1.6.2}$$

Recall the BSDE $(Y^{(\varepsilon)}, Z^{(\varepsilon)})$ from Definition 1.2.9 in Section 1.2.2. Since the representation Theorem 1.5.1 applies to $Z^{(\varepsilon)}$, we work with the version of $Z^{(\varepsilon)}$ given by

$$Z_t^{(\varepsilon)} = \mathbb{E}_t[\Phi(X_T)H_T^t + \int_t^T f^{(\varepsilon)}(s, X_s, Y_s^{(\varepsilon)}, Z_s^{(\varepsilon)})H_s^t ds] \quad \text{for all } t \in [0, T) \text{ almost surely.}$$

The following lemma decomposes the $\mathbf{L}_2$-regularity of Z into the $\mathbf{L}_2$-regularity of $Z^{(\varepsilon)}$ and terms controlled by ε.

Lemma 1.6.1. *Let $\beta \in (0, 1]$. Then there is a constant C such that for all $N > 1$*

$$\mathcal{E}(\pi_N^{(\beta)}) \leq \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \|Z_s - Z_{t_i}\|_2^2 ds \leq C \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \|Z_s^{(\varepsilon)} - Z_{t_i}^{(\varepsilon)}\|_2^2 ds + CN^{-2\gamma/\beta} + C\varepsilon^{2\gamma}\big(1 + \ln(N)\big).$$

Proof. In what follows, C may change in value from line to line.

Recall from Corollary 1.4.8 that $\int_0^T \|Z_s - Z_s^{(\varepsilon)}\|_2^2 ds \leq C\varepsilon^{2\gamma}$. Using the bound (1.4.14) with

$t = t_i$ for every $i \in \{0, \dots, N-1\}$ gives

$$\begin{aligned}
\sum_{i=0}^{N-1} \|Z_{t_i} - Z_{t_i}^{(\varepsilon)}\|_2^2 \Delta_i &\le C \sum_{i=0}^{N-1} \Big(\int_{t_i \vee (T-\varepsilon)}^{T} \frac{ds}{(T-s)^{1-\gamma}\sqrt{s-t_i}} \Big)^2 \Delta_i \\
&\le C \Big(\int_{t_{N-1} \vee (T-\varepsilon)}^{T} \frac{ds}{(T-s)^{1-\gamma}\sqrt{s-t_{N-1}}} \Big)^2 \Delta_{N-1} \\
&\quad + C \sum_{i=0}^{N-2} \Big(\int_{t_i \vee (T-\varepsilon)}^{T} \frac{ds}{(T-s)^{1-\gamma}} \Big)^2 \frac{\Delta_i}{t_{N-1} - t_i} \\
&\le C(\Delta_{N-1}^{\gamma-1/2})^2 \Delta_{N-1} + C\varepsilon^{2\gamma} + C\varepsilon^{2\gamma} \int_0^{t_{N-2}} \frac{ds}{t_{N-1} - s} \\
&\le CN^{-2\gamma/\beta} + C\varepsilon^{2\gamma}\big(1 + \ln(N)\big)
\end{aligned}$$

because $\Delta_{N-1} = TN^{-1/\beta}$. $\square$

In the following proposition, we obtain a convergence rate for the $\mathbf{L}_2$-regularity $\mathcal{E}(\pi)$ when π is a time grid of the form $\pi^{(\beta)}$. We do not use the results of Section 1.4. We will obtain a more precise convergence rate under stronger assumptions later in Theorem 1.6.13, but this first result will serve, for pedagogical purposes, to show that the results of Section 1.4 are useful for the estimation of the $\mathbf{L}_2$-regularity.

Proposition 1.6.2. *Let $0 < \beta < (2\gamma) \wedge \alpha$. There is a constant C depending only on L_f, C_M, θ_L, θ_c, β, C_f, $K^\alpha(\Phi)$, and T, but not on N, such that for all $N > 1$,*

$$\mathcal{E}(\pi_N^{(\beta)}) \le CN^{-1} + CN^{\big(1-(\alpha+\theta_L)\wedge 1\big)/\gamma - 1}.$$

Proof. In what follows, C may change in value from line to line.

To start with, assume $(\mathbf{A_{\partial f}})$ and $(\mathbf{A_{b\Phi}})$.

Recall the BSDEs $(y^{(\varepsilon)}, z^{(\varepsilon)})$ from Definition 1.2.9 and $(U^{(\varepsilon)}, V^{(\varepsilon)})$ from (1.2.14) in Section 1.2.2. It was stated in Lemma 1.2.14 that $z_t^{(\varepsilon)} = U_t^{(\varepsilon)}\sigma(t, X_t)$ and $(V_{j,t}^{(\varepsilon)})^\top = (\nabla z_{j,t}^{(\varepsilon)})^\top \sigma^{-1}(t, X_t) - U_t^{(\varepsilon)} \nabla_x \sigma_j(t, X_t)$
$m \times \mathbb{P} - a.e.$ for all $i \in \{1, \dots, q\}$, where $V_j^{(\varepsilon)}$ (resp. $\nabla z_j^{(\varepsilon)}$) is the j-th column of $V^{(\varepsilon)}$ (resp. $\nabla z^{(\varepsilon)}$), and σ_j is the j-th column of σ. In the proof of [GM10, Theorem 3.1], the authors show that for any i and $s \in [t_i, t_{i+1})$,

$$\|z_s^{(\varepsilon)} - z_{t_i}^{(\varepsilon)}\|_2 \le C \int_{t_i}^{s} \|a_r^{(\varepsilon)}\|_2 dr + C \int_{t_i}^{s} \|V_r^{(\varepsilon)}\|_2 dr + C\Delta_i^{1/2}. \tag{1.6.3}$$

Using $(\int_0^T \|a_r^{(\varepsilon)}\|_2 dr)^2 + \int_0^T \|V_r^{(\varepsilon)}\|_2^2 dr \le C\varepsilon^{-1+(\theta_L+\alpha)\wedge 1}$ from (1.2.15) in Lemma 1.2.14, it follows from Jensen's inequality that

$$\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \|z_s^{(\varepsilon)} - z_{t_i}^{(\varepsilon)}\|_2^2 ds \le \frac{C}{N} + \frac{C \max_{0 \le i \le N-1} \Delta_i}{\varepsilon^{1-(\theta_L+\alpha)\wedge 1}} \le \frac{C}{N} + \frac{C}{N\varepsilon^{1-(\theta_L+\alpha)\wedge 1}}$$

where $\max_i \Delta_i \leq CN^{-1}$ follows from (1.8.4) in Lemma 1.8.4. Combining this estimate with

$$\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \|z_s - z_{t_i}\|_2^2 ds \leq CN^{-1},$$

shown in [GM10, Theorem 1.3], $Z^{(\varepsilon)} = z + z^{(\varepsilon)}$, and, the results of Lemma 1.6.1, it follows that

$$\mathcal{E}(\pi_N^{(\beta)}) \leq \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \|Z_s - Z_{t_i}\|_2^2 ds \leq CN^{-2\gamma/\beta} + C\varepsilon^{2\gamma}\big(1 + \ln(N)\big) + \frac{C}{N\varepsilon^{1-\left((\theta_L+\alpha)\wedge 1\right)}} + \frac{C}{N}.$$

The proof under $(\mathbf{A_{\partial f}})$ and $(\mathbf{A_{b\Phi}})$ is completed by taking $\varepsilon = N^{-1/\gamma}$ and noticing that $2\gamma/\beta > 1$.

In order to prove the general result, recall the BSDE (Y_M, Z_M) from Corollary 1.4.9, and it satisfies the $\mathbf{L}_2$ regularity result of the proposition statement because its terminal condition satisfies $(\mathbf{A_{b\Phi}})$ and its driver satisfies $(\mathbf{A_{\partial f}})$. Moreover, [GM10, Lemma 3.1] yields $K^\alpha(\Phi_M) \leq K^\alpha(\Phi)$. Therefore, working with the version of Z_M given by

$$Z_{M,t} = \mathbb{E}_t[\Phi_M(X_T)H_T^t + \int_t^T f_M(r, X_r, Y_{M,r}, Z_{M,r})H_r^t dr] \quad \text{for all } t \in [0,T) \text{ almost surely,}$$

the result is extended to (Y, Z) using

$$\mathcal{E}(\pi_N^{(\beta)}) \leq \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \|Z_s - Z_{M,s}\|_2^2 ds + \sum_{i=0}^{N-1} \|Z_{t_i} - Z_{M,t_i}\|_2^2 \Delta_i + CN^{-1} + CN^{\left(1-(\alpha+\theta_L)\wedge 1\right)/\gamma - 1},$$

and letting $M \to \infty$ with Corollary 1.4.9. □

Remark. The proof method of Proposition 1.6.2 is in the spirit of [GM10, Theorem 3.1], where the authors make use of Lemma 1.2.14 to bound $\int_0^T \|V_r^{(\varepsilon)}\|_2^2 dr$ by $C\int_0^T \|a_r^{(\varepsilon)}\|_2^2 dr$. Since the driver f is uniformly Lipschitz (i.e. $\theta_L = 1$) in [GM10], there is an ε-uniform bound $\int_0^T \|a_r^{(\varepsilon)}\|_2^2 dr \leq C$, and it follows from Jensen's inequality that $\sum_i \int_{t_i}^{t_{i+1}} \|z_s^{(\varepsilon)} - z_{t_i}^{(\varepsilon)}\|_2^2 ds$ is bounded by $C \max_i \Delta_i$.

According to Proposition 1.6.2, the optimal rate of convergence CN^{-1} is obtained for $\alpha+\theta_L \geq 1$. In the case $\alpha = \theta_L$, this implies that $\alpha \geq 1/2$ is required. In the remainder of this section, we show that it is possible to slightly improve the rate of convergence under $(\mathbf{A_{exp\Phi}})$ by using a refined a priori estimates on $\|V_t^{(\varepsilon)}\|_2$. This a priori estimate will be very important in Section 1.7.

Proposition 1.6.3. *Suppose that $(\mathbf{A_{\partial f}})$ and $(\mathbf{A_{b\Phi}})$ are in force. There exists version of $V^{(\varepsilon)}$ and a constant C such that for any $\varepsilon \in (0,T]$ and severy $t \in [0,T)$, $\|V_t^{(\varepsilon)}\|_2 \leq C\phi(t,\varepsilon,\theta_L)$, where*

$$\phi(t,\varepsilon,\theta_L) := \|\Phi\|_\infty \int_t^{T-\varepsilon} \frac{dr}{(T-r)^{(3-\theta_L)/2}\sqrt{r-t}}. \tag{1.6.4}$$

Remark 1.6.4. The integral in (1.6.4) exists and is bounded by $C\varepsilon^{-(1-\theta_L)/2}(T-t)^{(\alpha-1)/2}$.

Proof. In what follows, C depends only on L_f, the bounds on b and σ and their partial derivatives, $\bar{\beta}$, C_M, θ_L, θ_c, C_f, and T, but not on ε, whose value may change from line to line.

For all $(t,x) \in [0,T) \times \mathbb{R}^d$, define the FBSDE

$$\left.\begin{array}{rcl} X_s^{(t,x)} & = & x + \int_t^s b(r, X_r^{(t,x)})\mathbf{1}_{(t,T]}(r)dr + \int_t^s \sigma(r, X_r^{(t,x)})\mathbf{1}_{(t,T]}(r)dW_r, \\ y_s^{(\varepsilon,t,x)} & = & \int_s^T F(r, X_r^{(t,x)}, y_r^{(\varepsilon,t,x)}, z_r^{(\varepsilon,t,x)})dr - \int_s^T z_r^{(\varepsilon,t,x)}dW_r \end{array}\right\} \tag{1.6.5}$$

where $F(t,x,y,z) = f^{(\varepsilon)}(t,x,u(t,x)+y, \nabla_x u(t,x)\sigma(t,x)+z)$. Note that the BSDE $(y^{(\varepsilon)}, z^{(\varepsilon)})$ from Section 1.2.2 is equal to $(y^{(\varepsilon,0,x_0)}, z^{(\varepsilon,0,x_0)})$ because (y,z) is equal to $\big(u(\cdot, X_\cdot), \nabla_x u(\cdot, X_\cdot)\sigma(\cdot, X_\cdot)\big)$ and X is equal to $X^{(0,x_0)}$. Since $f^{(\varepsilon)}(t,\cdot)$ is Lipschitz continuous for all $t \in [0,T]$, $F(t,\cdot)$ is also Lipschitz continuous for all $t \in [0,T)$: for all $(x,y,z), (x',y',z') \in \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^q$

$$\begin{aligned} |F(t,x,y,z) - F(t,x',y',z')| &\leq \frac{C\mathbf{1}_{[0,T-\varepsilon)}(t)}{\varepsilon^{(1-\theta_L)/2}}(|x-x'| + |y-y'| + |z-z'|) \\ &\quad + \frac{C\mathbf{1}_{[0,T-\varepsilon)}(t)}{\varepsilon^{(1-\theta_L)/2}}(|u(t,x) - u(t,x')| + |\nabla_x u(t,x) - \nabla_x u(t,x')|) \\ &\leq \frac{C\mathbf{1}_{[0,T-\varepsilon)}(t)}{\varepsilon^{(3-\theta_L)/2}}(|x-x'| + |y-y'| + |z-z'|) \end{aligned}$$

where the last inequality follows from the fact that $u(t,\cdot)$ and $\nabla_x u(t,\cdot)$ are differentiable, and their derivatives are bounded by $C(T-t)^{-1/2}$ and $C(T-t)^{-1}$ respectively, Lemma 1.2.10.

Let $H_r^{(t,x,s)} := \frac{\mathbf{1}_{(t,T]}(s)}{r-s}(\int_s^r \sigma^{-1}(r, X_r^{(t,x)}) D_s X_r^{(t,x)})^\top dW_r)^\top$ where $D_s X^{(t,x)}$ is the Malliavin derivative of $X^{(t,x)}$ evaluated at time s, as defined in Section 1.2.1. One can show analogously to Theorem 1.5.1, with $X^{(t,x)}$ replacing X and $H_r^{(t,x,s)}$ replacing H_r^s, that there is a predictable version $\mathcal{Z}$ of $z^{(\varepsilon,t,x)}$ such that

$$\mathcal{Z}_s = \mathbb{E}_s[\int_s^T F(r, X_r^{(t,x)}, y_r^{(\varepsilon,t,x)}, z_r^{(\varepsilon,t,x)})H_r^{(t,x,s)}dr] \quad \text{for all } s \in [t,T) \text{ almost surely;} \tag{1.6.6}$$

we work with this version of $z^{(\varepsilon,t,x)}$ from hereon. Moreover, there is a continuous function $z^{(\varepsilon)} : [0,T) \times \mathbb{R}^d \to \mathbb{R}^q$ given by

$$z^{(\varepsilon)}(t,x) = \mathbb{E}[\int_t^T F(r, X_r^{(t,x)}, y_r^{(\varepsilon,t,x)}, z_r^{(\varepsilon,t,x)})H_r^{(t,x,t)}dr] \tag{1.6.7}$$

for all $(t,x) \in [0,T) \times \mathbb{R}^d$ and $z_t^{(\varepsilon)} = z^{(\varepsilon)}(t, X_t)$ for all $t \in [0,T)$ almost surely. The proof of this can be found in [MZ02, Theorem 4.2]; the conditions of [MZ02, Theorem 3.1] are satisfied because the terminal condition of the BSDE $(y^{(\varepsilon)}, z^{(\varepsilon)})$ is zero.

Fix $s \in [t,T)$. Using the representation (1.6.7) of $z^{(\varepsilon,t,x)}$, it follows that

$$\begin{aligned} \|z_s^{(\varepsilon,t,x_1)} - z_s^{(\varepsilon,t,x_2)}\|_2 \leq \|\mathbb{E}_s[&\int_s^T F(r, X_r^{(t,x_1)}, y_r^{(\varepsilon,t,x_1)}, z_r^{(\varepsilon,t,x_1)})H_r^{(t,x_1,s)}dr] \\ &- \mathbb{E}_s[\int_s^T F(r, X_r^{(t,x_2)}, y_r^{(\varepsilon,t,x_2)}, z_r^{(\varepsilon,t,x_2)})H_r^{(t,x_1,s)}dr]\|_2 \\ &+ \|\mathbb{E}_s[\int_s^T F(r, X_r^{(t,x_2)}, y_r^{(\varepsilon,t,x_2)}, z_r^{(\varepsilon,t,x_2)})(H_r^{(t,x_1,s)} - H_r^{(t,x_2,s)})dr]\|_2 \\ =:\ & \mathcal{A}_1 + \mathcal{A}_2. \end{aligned}$$

We start with an estimate for $\mathcal{A}_2$. Using Fubini's theorem, Lemma 1.8.1, and the Cauchy-Schwarz inequality, it follows that

$$\mathcal{A}_2 \leq \int_s^T \|F(r, X_r^{(t,x_2)}, y_r^{(\varepsilon,t,x_2)}, z_r^{(\varepsilon,t,x_2)})\|_4 \|H_r^{(t,x_1,s)} - H_r^{(t,x_2,s)}\|_4 dr$$

Using the same techniques as in the proof of Lemma 1.4.1, but replacing the Itô isometry with the Burkholder-Davis-Gundy inequality,

$$\begin{aligned}\|H_r^{(t,x_1,s)} - H_r^{(t,x_2,s)}\|_4 \leq C_4 &\frac{\|\sigma^{-1}(s, X_s^{(t,x_1)}) - \sigma^{-1}(s, X_s^{(t,x_2)})\|_{\mathcal{S}^8} \mathbb{E}[\sup_{s\leq u\leq T} |D_s X_u^{(t,x_1)}|^8]^{1/8}}{\sqrt{r-s}} \\ &+ C_4 \frac{\|\sigma^{-1}\|_\infty \mathbb{E}[\sup_{s\leq u\leq T} |D_s X_u^{(t,x_1)} - D_s X_u^{(t,x_2)}|^8]^{1/8}}{\sqrt{r-t}}.\end{aligned}$$

where C_4 is the constant coming from the BDG inequality. The function $\sigma^{-1}(t,\cdot)$ is Lipschitz continuous uniformly in t with Lipschitz constant as given in Lemma 1.1.4. $X^{(t,x)}$ solves a linear BSDE for all (t,x), and, as for $D_s X$ in Lemma 1.2.6, $D_s X^{(t,x)}$ also solves a linear SDE for all (t,x). One can then use the proof method of [RY99, Theorem IX.2.4] (essentially using Gronwall's inequality) to show that

$$\|X_s^{(t,x_1)} - X_s^{(t,x_2)}\|_{\mathcal{S}^8} + \mathbb{E}[\sup_{s\leq u\leq T} |D_s X_u^{(t,x_1)} - D_s X_u^{(t,x_2)}|^8]^{1/8} \leq C|x_1 - x_2| \tag{1.6.8}$$

for all $s \in [t,T)$. Moreover, $\mathbb{E}[\sup_{s\leq u\leq T} |D_s X_u^{(t,x_1)}|^8]^{1/8} \leq C$. Using (1.6.8), it follows that $\|H_r^{(t,x_1,t)} - H_r^{(t,x_2,t)}\|_4 \leq C|x_1 - x_2|/\sqrt{r-t}$.

In order to find a bound for $\|F(r, X_r^{(t,x_2)}, y_r^{(\varepsilon,t,x_2)}, z_r^{(\varepsilon,t,x_2)})\|_4$, we show that $y_r^{(\varepsilon,t,x_2)}$ and $z_r^{(\varepsilon,t,x_2)}$ are in $\mathbf{L}_4(\mathcal{F}_r)$. Once this has been shown, we take advantage of the local Lipschitz continuity and boundedness (1.1.3) of f, and the uniform bounds on u and its partial derivatives from Lemma 1.2.10, in order to show that

$$\begin{aligned}|F(r, X_r^{(t,x_2)}, 0, 0)| &\leq |f(r, X_r^{(t,x_2)}, 0, 0)| + L_f \frac{|u(r, X_r^{(t,x_2)})| + \|\sigma\|_\infty |\nabla_x u(r, X_r^{(t,x_2)})|}{(T-r)^{(1-\theta_L)/2}} \\ &\leq \frac{C_f}{(T-r)^{1-\theta_c}} + \frac{C\|\Phi\|_\infty}{(T-r)^{(3-\theta_L)/2}} \leq \frac{C\|\Phi\|_\infty}{(T-r)^{(3-\theta_L)/2}}\end{aligned} \tag{1.6.9}$$

It follow from the triangle inequality, the local Lipschitz continuity (1.1.3) and the inequality (1.6.9) that

$$\begin{aligned}\|F(r, X_r^{(t,x_2)}, y_r^{(\varepsilon,t,x_2)}, z_r^{(\varepsilon,t,x_2)})\|_4 &\leq \|F(r, X_r^{(t,x_2)}, 0, 0)\|_4 + L_f \frac{\|y_r^{(\varepsilon,t,x_2)}\|_4 + \|z_r^{(\varepsilon,t,x_2)}\|_4}{(T-r)^{(1-\theta_L)/2}} \\ &\leq \frac{C\|\Phi\|_\infty}{(T-r)^{(3-\theta_L)/2}} + L_f \frac{\|y_r^{(\varepsilon,t,x_2)}\|_4 + \|z_r^{(\varepsilon,t,x_2)}\|_4}{(T-r)^{(1-\theta_L)/2}}.\end{aligned} \tag{1.6.10}$$

To show that show that $y_r^{(\varepsilon,t,x_2)}$ and $z_r^{(\varepsilon,t,x_2)}$ are in $\mathbf{L}_4(\mathcal{F}_r)$, apply Lemma 1.4.5 with $(Y_1, Z_1) = (0,0)$

and $(Y_2, Z_2) = (y^{(\varepsilon,t,x_2)}, z^{(\varepsilon,t,x_2)})$, combined with inequality (1.6.9) to obtain that

$$\left.\begin{aligned} |y_r^{(\varepsilon,t,x_2)}| &\leq C \textstyle\int_r^{T-\varepsilon} \mathbb{E}_r[|F(u, X_u^{(t,x_2)}, 0, 0)|^2]^{1/2} du \leq C\|\Phi\|_\infty \textstyle\int_r^{T-\varepsilon} (T-u)^{(\theta_L - 3)/2} du, \\ |z_r^{(\varepsilon,t,x_2)}| &\leq C \textstyle\int_r^{T-\varepsilon} \mathbb{E}_r[|F(u, X_u^{(t,x_2)}, 0, 0)|^2]^{1/2} (u-r)^{-1/2} du \\ &\qquad\qquad \leq C\|\Phi\|_\infty \textstyle\int_r^{T-\varepsilon} (T-u)^{(\theta_L-3)/2} (u-r)^{-1/2} du \end{aligned}\right\} \tag{1.6.11}$$

for all $r \in [t, T)$, where we have used (1.6.9). Therefore,

$$\|F(r, X_r^{(t,x_2)}, y_r^{(\varepsilon,t,x_2)}, z_r^{(\varepsilon,t,x_2)})\|_4 \leq \frac{C\|\Phi\|_\infty}{(T-r)^{(3-\theta_L)/2}} + C\|\Phi\|_\infty \frac{\int_r^{T-\varepsilon} (T-u)^{(\theta_L-3)/2} (u-r)^{-1/2} du}{(T-r)^{(1-\theta_L)/2}}.$$

Combining the above estimates and Lemma 1.8.5, it follows that

$$\begin{aligned} \mathcal{A}_2 &\leq C\|\Phi\|_\infty |x_1 - x_2| \{ \int_s^{T-\varepsilon} \frac{dr}{(T-r)^{(3-\theta_L)/2}\sqrt{r-s}} + \int_s^{T-\varepsilon} \frac{\int_r^{T-\varepsilon}(T-u)^{(3-\theta_L)/2}(u-r)^{-1/2} du}{(T-r)^{(1-\theta_L)/2}\sqrt{r-s}} dr \} \\ &= C\|\Phi\|_\infty |x_1 - x_2| \{ \int_s^{T-\varepsilon} \frac{dr}{(T-r)^{(3-\theta_L)/2}\sqrt{r-s}} + \int_s^{T-\varepsilon} \frac{\int_s^u (u-r)^{\theta_L/2-1}(r-s)^{-1/2} dr}{(T-u)^{(3-\theta_L)/2}} du \} \\ &\leq C\|\Phi\|_\infty |x_1 - x_2| \int_s^{T-\varepsilon} \frac{dr}{(T-r)^{((3-\theta_L)/2}\sqrt{r-s}}. \end{aligned}$$

Now, we estimate $\mathcal{A}_1$. Using the conditional Fubini's theorem, Lemma 1.8.1, and the Cauchy-Schwarz inequality,

$$\mathcal{A}_1 \leq \| \int_s^{T-\varepsilon} \mathbb{E}_s[|F(r, X_r^{(t,x_1)}, y_r^{(\varepsilon,t,x_1)}, z_r^{(\varepsilon,t,x_1)}) - F(r, X_r^{(t,x_2)}, y_r^{(\varepsilon,t,x_2)}, z_r^{(\varepsilon,t,x_2)})|^2]^{1/2} \mathbb{E}_s[|H_r^{(t,x_1,s)}|^2]^{1/2} dr\|_2$$

Analogously to Lemma 1.4.1, $\mathbb{E}_s[|H_r^{(t,x_1,s)}|^2]^{1/2} \leq C(r-t)^{-1/2}$, therefore Minkowski's inequality implies

$$\mathcal{A}_1 \leq C \int_s^{T-\varepsilon} \frac{\|F(r, X_r^{(t,x_1)}, y_r^{(\varepsilon,t,x_1)}, z_r^{(\varepsilon,t,x_1)}) - F(r, X_r^{(t,x_2)}, y_r^{(\varepsilon,t,x_2)}, z_r^{(\varepsilon,t,x_2)})\|_2}{\sqrt{r-s}} dr.$$

Define $(Y^{(\varepsilon,t,x)}, Z^{(\varepsilon,t,x)}) := (u(\cdot, X_\cdot^{(t,x)}) + y^{(\varepsilon,t,x)}, \nabla_x u(\cdot, X_\cdot^{(t,x)})\sigma(\cdot, X_\cdot^{(t,x)}) + z^{(\varepsilon,t,x)})$. By applying

Minkowski's inequality and the Lipschitz continuity of $f^{(\varepsilon)}(r,\cdot)$, $\mathcal{A}_1$ is bounded by

$$
\begin{aligned}
&C\int_s^{T-\varepsilon} \frac{\|X_r^{(t,x_1)}-X_r^{(t,x_2)}\|_2+\|Y_r^{(\varepsilon,t,x_1)}-Y_r^{(\varepsilon,t,x_2)}\|_2+\|Z_r^{(\varepsilon,t,x_1)}-Z_r^{(\varepsilon,t,x_2)}\|_2}{(T-r)^{(1-\theta_L)/2}\sqrt{r-t}}dr \\
&\leq C\int_s^{T-\varepsilon} \frac{\|X_r^{(t,x_1)}-X_r^{(t,x_2)}\|_2+\|\sigma\|_\infty\|u(r,X_r^{(t,x_1)})-u(r,X_r^{(t,x_2)})\|_2}{(T-r)^{(1-\theta_L)/2}\sqrt{r-t}}dr \\
&\quad + C\int_s^{T-\varepsilon} \frac{\|\nabla_x\sigma\|_\infty\|\nabla_x u(r,X_r^{(t,x_1)})-\nabla_x u(r,X_r^{(t,x_2)})\|_2}{(T-r)^{(1-\theta_L)/2}\sqrt{r-t}}dr \\
&\quad + C\int_s^{T-\varepsilon} \frac{\|y_r^{(\varepsilon,t,x_1)}-y_r^{(\varepsilon,t,x_2)}\|_2+\|z_r^{(\varepsilon,t,x_1)}-z_r^{(\varepsilon,t,x_2)}\|_2}{(T-r)^{(1-\theta_L)/2}\sqrt{r-t}}dr
\end{aligned}
$$

Using the differentiability of $u(s,\cdot)$ and the boundedness of its partial derivatives from Lemma 1.2.10, and the bound $\|X_r^{(t,x_1)}-X_r^{(t,x_2)}\|_4 \leq C|x_1-x_2|$ obtained from [RY99, Theorem IX.2.4], it follows that

$$
\begin{aligned}
&\|u(r,X_r^{(t,x_1)})-u(r,X_r^{(t,x_2)})\|_2 \leq \|X_r^{(t,x_1)}-X_r^{(t,x_2)}\|_4\|\mathcal{R}(u,r,X_r^{(t,x_1)},X_r^{(t,x_2)})\|_4 \leq \frac{C\|\Phi\|_\infty|x_1-x_2|}{(T-r)^{-1/2}}, \\
&\|\nabla_x u(r,X_r^{(t,x_1)})-\nabla_x u(r,X_r^{(t,x_2)})\|_2 \leq \|X_r^{(t,x_1)}-X_r^{(t,x_2)}\|_4\|\mathcal{R}(\nabla_x u,r,X_s^{(t,x_1)},X_s^{(t,x_2)})\|_4 \leq \frac{C\|\Phi\|_\infty|x_1-x_2|}{(T-r)^{-1}}
\end{aligned}
$$

for all $r\in[t,T)$, where $\mathcal{R}(g,r,x,x')$ is the remainder from the Taylor expansion of $g(r,x)-g(r,x')$. Therefore, denoting by $\Theta_r := \|y_r^{(\varepsilon,t,x_1)}-y_r^{(\varepsilon,t,x_2)}\|_2+\|z_r^{(\varepsilon,t,x_1)}-z_r^{(\varepsilon,t,x_2)}\|_2$, the final bound on $\mathcal{A}_1$ is

$$
\mathcal{A}_1 \leq C\|\Phi\|_\infty|x_1-x_2|\int_s^{T-\varepsilon}\frac{dr}{(T-r)^{(3-\theta_L)/2}\sqrt{r-t}}+C\int_s^{T-\varepsilon}\frac{\Theta_r}{(T-r)^{(1-\theta_L)/2}\sqrt{r-t}}dr.
$$

From the bounds on $\mathcal{A}_1$ and $\mathcal{A}_2$, it follows that

$$
\|z_s^{(\varepsilon,t,x_1)}-z_s^{(\varepsilon,t,x_2)}\|_2 \leq C\|\Phi\|_\infty|x_1-x_2|\int_s^{T-\varepsilon}\frac{dr}{(T-r)^{(3-\theta_L)/2}\sqrt{r-s}}+C\int_s^{T-\varepsilon}\frac{\Theta_r}{(T-r)^{(1-\theta_L)/2}\sqrt{r-s}}dr
$$

Since $y_s^{(\varepsilon,t,x)}=\mathbb{E}_s[\int_s^T f^{(\varepsilon)}(r,X_r^{(t,x)},Y_r^{(\varepsilon,t,x)},Z_r^{(\varepsilon,t,x)})dr]$, one can show analogously that

$$
\|y_s^{(\varepsilon,t,x_1)}-y_s^{(\varepsilon,t,x_2)}\|_2 \leq C\|\Phi\|_\infty\int_s^{T-\varepsilon}\frac{dr}{(T-r)^{(3-\theta_L)/2}}+C\int_s^{T-\varepsilon}\frac{\Theta_r}{(T-r)^{(1-\theta_L)/2}}dr,
$$

whence it follows that

$$
\Theta_s \leq C\|\Phi\|_\infty|x_1-x_2|\int_s^{T-\varepsilon}\frac{dr}{(T-r)^{(3-\theta_L)/2}\sqrt{r-s}}+C\int_s^{T-\varepsilon}\frac{\Theta_r}{(T-r)^{(1-\theta_L)/2}\sqrt{r-s}}dr
$$

for all $s\in[t,T-\varepsilon)$. Applying Lemma 1.8.6 with $w_r = C\|\Phi\|_\infty|x_1-x_2|\int_r^{T-\varepsilon}(T-u)^{(\theta_L-3)/2}(u-$

$r)^{-1/2}du$ and $u_r = \Theta_r$, it follows that

$$\begin{aligned}
\Theta_s &\leq C\|\Phi\|_\infty |x_1 - x_2| \int_s^{T-\varepsilon} \frac{dr}{(T-r)^{(3-\theta_L)/2}\sqrt{r-s}} \\
&\quad + C\|\Phi\|_\infty |x_1 - x_2| \int_s^{T-\varepsilon} \frac{\int_r^{T-\varepsilon}(T-u)^{(\theta_L-3)/2}(u-r)^{-1/2}du}{(T-r)^{(1-\theta_L)/2}\sqrt{r-s}} dr + C\int_s^{T-\varepsilon} \frac{\Theta_r}{(T-r)^{(1-\theta_L)/2}} dr \\
&= C\|\Phi\|_\infty |x_1 - x_2| \int_s^{T-\varepsilon} \frac{dr}{(T-r)^{(3-\theta_L)/2}\sqrt{r-s}} \\
&\quad + C\|\Phi\|_\infty |x_1 - x_2| \int_s^{T-\varepsilon} \frac{\int_s^{u}(u-r)^{\theta_L/2-1}(r-s)^{-1/2}dr}{(T-u)^{(3-\theta_L)/2}} du + C\int_s^{T-\varepsilon} \frac{\Theta_r}{(T-r)^{(1-\theta_L)/2}} dr \\
&\leq C\|\Phi\|_\infty |x_1 - x_2| \int_s^{T-\varepsilon} \frac{dr}{(T-r)^{(3-\theta_L)/2}\sqrt{r-s}} + C\int_s^{T-\varepsilon} \frac{\Theta_r}{(T-r)^{(1-\theta_L)/2}} dr,
\end{aligned}$$

where we have used Lemma 1.8.4 to bound the integral $\int_s^u (u-r)^{\theta_L/2-1}(r-s)^{-1/2}dr$ by $C(u-s)^{(\theta_L-1)s/2}$. Then, applying Lemma 1.8.7 to bound the integral $\int_s^{T-\varepsilon}\Theta_r(T-r)^{(\theta_L-1)/2}dr$, final bound on $\|z_s^{(\varepsilon,t,x_1)} - z_s^{(\varepsilon,t,x_2)}\|_2$ for all $t \in [0,T)$, $(x_1,x_2) \in (\mathbb{R}^d)^2$, and $s \in [t,T)$ is

$$\begin{aligned}
\|z_s^{(\varepsilon,t,x_1)} - z_s^{(\varepsilon,t,x_2)}\|_2 &\leq C\|\Phi\|_\infty |x_1 - x_2| \int_s^{T-\varepsilon} \frac{dr}{(T-r)^{(3-\theta_L)/2}\sqrt{r-s}} \\
&\quad + C\|\Phi\|_\infty |x_1 - x_2| \int_s^{T-\varepsilon} \frac{\int_r^{T-\varepsilon}(T-u)^{(\theta_L-3)/2}(u-r)^{-1/2}du}{(T-r)^{(1-\theta_L)/2}} dr \\
&\leq C\|\Phi\|_\infty |x_1 - x_2| \int_s^{T-\varepsilon} (T-r)^{(\theta_L-3)/2}(r-s)^{-1/2} dr
\end{aligned} \tag{1.6.12}$$

In particular, this implies that the function $z^{(\varepsilon)}(t,\cdot)$ is Lipschitz continuous with Lipschitz constant $C\|\Phi\|_\infty \int_t^{T-\varepsilon}(T-r)^{(\theta_L-3)/2}(r-s)^{-1/2}dr$. Using a standard mollification procedure, we can assume that $z^{(\varepsilon)}(t,\cdot)$ is in fact continuously differentiable with partial derivatives bounded by $C\|\Phi\|_\infty \int_t^{T-\varepsilon}(T-r)^{(\theta_L-3)/2}(r-s)^{-1/2}dr$. Hence, using the chain rule Lemma 1.2.1, the Malliavin derivative of $z_t^{(\varepsilon)}$ can be expressed as

$$D_s z_t^{(\varepsilon)} = (D_s X_t)^\top \nabla_x z^{(\varepsilon)}(t, X_t)$$

Recalling from Lemma 1.2.13 that $\nabla z_t^{(\varepsilon)} = \left(\sigma^{-1}(s,X_s)\nabla X_s\right)^\top D_s z_t^{(\varepsilon)}$, it follows from Lemma 1.2.7 that $\nabla z_t^{(\varepsilon)} = (\nabla X_t)^\top \nabla_x z^{(\varepsilon)}(t,X_t)$. This implies that $\|\nabla z_t^{(\varepsilon)}\|_2 \leq \phi(t,\varepsilon,\theta_L)$, thanks to the uniform bound on on $\nabla_x z(t,\cdot)$ and the bound $\|\nabla X\|_{\mathcal{S}^2} \leq C$. Now, using Lemma 1.2.14,

$$\|\sup_{s\leq r<T} U_r^{(\varepsilon)}\|_2 \leq C\int_s^{T-\varepsilon} \|a_r^{(\varepsilon)}\|_2 dr \leq C\int_s^{T-\varepsilon} \frac{dr}{(T-r)^{(3-\alpha)/2}} \leq \phi(s,\varepsilon,\theta_L),$$

and $(V_{j,t}^{(\varepsilon)})^\top = (\nabla z_{j,t}^{(\varepsilon)})^\top \sigma^{-1}(t,X_t) - U_t^{(\varepsilon)}\nabla_x \sigma_j(t,X_t)$ from Lemma 1.2.14, we conclude that $\|V_t^{(\varepsilon)}\|_2 \leq \phi(s,\varepsilon,\theta_L)$ as required. □

In order to make use of Proposition 1.6.3, it is is necessary decompose the $\mathbf{L}_2$-regularity of Z into the $\mathbf{L}_2$-regularity of an intermediate process Z_M, which satisfies the hypotheses of Proposition 1.6.3 and a small error term.

Lemma 1.6.5. *Assume that* $(\mathbf{A_{exp\Phi}})$ *is in force. Recall the BSDE* (Y_M, Z_M) *defined in Corollary 1.4.9. Take the version of* Z_M *satisfying* $Z_{M,t} = \mathbb{E}_t[\Phi_M(X_T)H_T^t + \int_t^T f_M(s, X_s, Y_{M,s}, Z_{M,s})H_s^t ds]$ *for all* $t \in [0,T)$ *almost surely. For* $M = 2\log(N)$ *and* R_t *sufficiently large, there is a constant* C *depending only on* L_f*,* C_M*,* θ_L*, and* T*, but not on* N*, such that for all* $N > 1$

$$\mathcal{E}(\pi_N^{(\beta)}) \le \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \|Z_s - Z_{t_i}\|_2^2 ds \le C \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \|Z_{M,s} - Z_{M,t_i}\|_2^2 ds + CN^{-1}.$$

Proof. Recall the BSDE $(\bar{Y}_M, \bar{Z}_M)$ from Corollary 1.4.10, and take the version of $\bar{Z}_M$ satisfying $\bar{Z}_{M,t} = \mathbb{E}_t[\Phi_M(X_T)H_T^t + \int_t^T f_M(s, X_s, \bar{Y}_{M,s}, \bar{Z}_{M,s})H_s^t ds]$ for all $t \in [0,T)$ almost surely. The corollary also states that there is a constant C such that

$$\int_0^T \|Z_s - \bar{Z}_{M,s}\|_2^2 ds \le C\|\Phi(X_T) - \Phi_M(X_T)\|_2^2, \qquad \|Z_{t_i} - \bar{Z}_{M,t_i}\|_2 \le \frac{C\|\Phi(X_T) - \Phi_M(X_T)\|_2}{\sqrt{T - t_i}} \ \forall t_i \in \pi.$$

It follows from Markov's exponential inequality that

$$\|\Phi(X_T) - \Phi_M(X_T)\|_2^2 = \int_{M^2}^{\infty} \mathbb{P}(\|\Phi(X_T)\|_2^2 \ge x) dx \le 2C_\xi \int_{M^4}^{\infty} e^{-\sqrt{x}} dx = 2C_\xi(1 + M^4)e^{-M^4}.$$

Taking $M = (1+\delta)^{1/4}\log(N)^{1/4}$ for $\delta \in (0,1]$ sufficiently large so that $N^{-\delta}\log(N) \le 1$, it follows that there is a (possibly different) constant C depending such that

$$\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \|Z_s - \bar{Z}_{M,s}\|_2^2 ds + \sum_{i=0}^{N-1} \|Z_{t_i} - \bar{Z}_{M,t_i}\|_2^2 \Delta_i \le CN^{-1}.$$

The triangle inequality yields the error decomposition

$$\begin{aligned}\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \|Z_s - Z_{t_i}\|_2^2 ds \le{}& \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \|Z_s - \bar{Z}_{M,s}\|_2^2 ds + \sum_{i=0}^{N-1} \|Z_{t_i} - \bar{Z}_{M,t_i}\|_2^2 \Delta_i \\ &+ \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \|Z_{M,s} - \bar{Z}_{M,s}\|_2^2 ds + \sum_{i=0}^{N-1} \|Z_{M,t_i} - \bar{Z}_{M,t_i}\|_2^2 \Delta_i \\ &+ \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \|Z_{M,s} - Z_{M,t_i}\|_2^2 ds.\end{aligned}$$

Applying Proposition 1.4.2 with $(Y_1, Z_1) = (\bar{Y}_M, \bar{Z}_M)$ and $(Y_2, Z_2) = (Y_M, Z_M)$, it follows that there is a (possibly different) finite $C \ge 0$ depending only on L_f, θ_L C_M, and T such that

$$\begin{aligned}\int_0^T \|Z_{M,t} - \bar{Z}_{M,t}\|_2^2 dt &\le C \int_0^T \|f(t, X_t, \bar{Y}_{M,t}, \bar{Z}_{M,t}) - f_M(t, X_t, \bar{Y}_{M,t}, \bar{Z}_{M,t}\|_2^2 dt, \\ \|Z_{M,t_i} - \bar{Z}_{M,t_i}\|_2 &\le C \int_{t_i}^T \frac{\|f(t, X_t, \bar{Y}_{M,t}, \bar{Z}_{M,t}) - f_M(t, X_t, \bar{Y}_{M,t}, \bar{Z}_{M,t})\|_2}{\sqrt{t - t_i}} dt \quad \forall t_i \in \pi.\end{aligned}$$

Using the definition of the function $\phi_{R,M}$, in Definition 1.1.1, and the definition of f_M, it follows that

$$
\begin{aligned}
&|f(t, X_t, \bar{Y}_{M,t}, \bar{Z}_{M,t}) - f_M(t, X_t, \bar{Y}_{M,t}, \bar{Z}_{M,t})| \\
&\leq \int_{\mathbb{R}^d \times \mathbb{R} \times (\mathbb{R}^q)^\top} |f(t, X_t, \bar{Y}_{M,t}, \bar{Z}_{M,t}) - f(t, X_t - x, \bar{Y}_{M,t} - y, \bar{Z}_{M,t} - z)| \phi_{R_t,M}(x,y,z) d(x,y,z) \\
&\leq \frac{L_f}{(T-t)^{(1-\theta_L)/2}} \int_{\{|x|^2+|y|^2+|z|^2 \leq R_t^{-2} M^{-2}\}} (|x|+|y|+|z|) \phi_{R_t,M}(x,y,z) d(x,y,z). \leq \frac{3L_f}{(T-t)^{(1-\theta_L)/2} R_t M}
\end{aligned}
$$

Setting $M = (1+\delta)^{1/4} \log(N)^{1/4}$ as above, and the parameter R_t to be equal to $L_f N^{1/2} M^{-1} (T - t)^{(\theta_L - 1)/2}$ ensures that $\int_0^T \|Z_{M,t} - \bar{Z}_{M,t}\|_2^2 dt \leq CN^{-1}$ and $\|Z_{M,t_i} - \bar{Z}_{M,t_i}\|_2 \leq CN^{-1/2}(T-t_i)^{1/2}$, whence the result follows. □

We now come to the main result of this section, where a convergence rate for the $\mathbf{L}_2$ regularity is computed with the aid of Proposition 1.6.3.

Theorem 1.6.6. *Assume that* $(\mathbf{A_{exp\Phi}})$ *is in force. Suppose that* $0 < \beta < (2\gamma) \wedge \alpha$. *There is a constant* C *such that for all* N,

$$
\mathcal{E}(\pi_N^{(\beta)}) \leq \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \|Z_s - Z_{t_i}\|_2^2 ds \leq CN^{-1} + CN^{-2+(3\theta_L/4-1)(1+\delta_N)/(2\gamma)} (\ln(N) \vee 1) \tag{1.6.13}
$$

where $\delta_N \geq \mathbf{1}_{[3,\infty)}(N) \ln\ln(N)/\ln(N)$.

Proof. In what follows, C may change from line to line.

To start with, we assume that $(\mathbf{A_{\partial f}})$ and $(\mathbf{A_{b\Phi}})$ are in force. Recall (1.6.3). From the bounds $\|a_r^{(\varepsilon)}\|_2 \leq C(T-r)^{(\alpha+\theta_L-3)/2}$ in the proof on Lemma 1.2.14 the first sum $\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} (\int_{t_i}^t \|a_r^{(\varepsilon)}\|_2 dr)^2 dt$ is bounded above by

$$
\begin{aligned}
&C \sum_{i=0}^{N-2} \int_{t_i}^{t_{i+1}} \frac{\left(\int_{t_i}^t \frac{dr}{(t-r)^{(1-\theta_L)/2}}\right)^2}{(T-t)^{2-\alpha}} dt + C \int_{t_{N-1}}^T \frac{\left(\int_{t_{N-1}}^t \frac{dr}{(t-r)^{1-\theta_L/2}}\right)^2}{(T-t)^{1-\alpha}} dt \\
&\leq C \sum_{i=0}^{N-2} \int_{t_i}^{t_{i+1}} \frac{(t-t_i)^{1+\theta_L}}{(T-t)^{2-\alpha}} dt + C \int_{t_{N-1}}^T \frac{(t-t_{N-1})^{\theta_L}}{(T-t)^{1-\alpha}} dt \\
&\leq C \sum_{i=0}^{N-2} \frac{\Delta_i^{1+\theta_L}}{(T-t_{i+1})^{1-\beta}} \int_{t_i}^{t_{i+1}} \frac{dt}{(T-t)^{1+\beta-\alpha}} + C\Delta_{N-1}^{\theta_L+\alpha}.
\end{aligned} \tag{1.6.14}
$$

Using (1.8.5) from Lemma 1.8.4, $\frac{\Delta_i}{\Delta_{i+1}} \leq C$ for $i < N-1$, which, combined with (1.8.4), yields

$$
\max_{0 \leq i \leq N-2} \frac{\Delta_i^{1+\theta_L}}{(T-t_{i+1})^{1-\beta}} \leq C \max_{0 \leq i \leq N-1} \frac{\Delta_i^{1+\theta_L}}{(T-t_i)^{1-\beta}} < CN^{-1-\theta_L}.
$$

Additionally, $\beta < \alpha$ implies that $\Delta_{N-1}^{\alpha+\theta_L} = CN^{-(\alpha+\theta_L)/\beta} \leq CN^{-1}$. Substituting these results into (1.6.14) gives

$$
\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} (\int_{t_i}^t \|a_r^{(\varepsilon)}\|_2 dr)^2 dt \leq CN^{-1}. \tag{1.6.15}
$$

The refined estimates $\|V_r^{(\varepsilon)}\|_2 \leq C\phi(r, \varepsilon, \theta_L)\ dr - a.e.$ from Proposition 1.6.3 are used to bound

$\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}(\int_{t_i}^{t}\|V_r^{(\varepsilon)}\|_2 dr)^2 dt$. Using Lemma 1.8.5 and Jensen's inequality,

$$
\begin{aligned}
&\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\Big(\int_{t_i}^{t}\phi(r,\varepsilon,\theta_L)dr\Big)^2 dt\\
&= C\|\Phi\|_\infty\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\Big(\int_{t_i}^{t}\Big\{\int_r^{T-\varepsilon}\frac{du}{(T-u)^{\frac{3-\theta_L}{2}}\sqrt{u-r}}\Big\}dr\Big)^2 dt\\
&\le C\|\Phi\|_\infty\frac{1}{\varepsilon^{1-3\theta_L/4}}\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\Big(\int_{t_i}^{t}\Big\{\int_r^{T-\varepsilon}\frac{du}{(T-u)^{1-\theta_L/8}\sqrt{u-r}}\Big\}dr\Big)^2 dt\\
&\le C\|\Phi\|_\infty\frac{1}{\varepsilon^{1-3\theta_L/4}}\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\Big(\int_{t_i}^{t}\frac{dr}{(T-r)^{\left(1-\theta_L/4\right)/2}}\Big)^2 dt\\
&\le \frac{C\|\Phi\|_\infty(\max_{0\le i\le N-1}\Delta_i)^2}{\varepsilon^{1-3\theta_L/4}}\int_0^T\frac{dr}{(T-t)^{1-\theta_L/4}}dt \le \frac{C\|\Phi\|_\infty N^{-2}}{\varepsilon^{1-3\theta_L/4}}
\end{aligned}
\tag{1.6.16}
$$

where we have used (1.8.4) in Lemma 1.8.4 for the bound $\max_{0\le i\le N-1}\Delta_i \le CN^{-1}$. Substituting (1.6.15) and (1.6.16) into (1.6.3) finally yields

$$
\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\|z_s^{(\varepsilon)}-z_{t_i}^{(\varepsilon)}\|_2^2 ds \le CN^{-1}+\frac{C\|\Phi\|_\infty N^{-2}}{\varepsilon^{1-3\theta_L/4}}. \tag{1.6.17}
$$

Then, using $Z^{(\varepsilon)}=z^{(\varepsilon)}+z$, Lemma 1.6.1, and $\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\|z_s-z_{t_i}\|_2^2 ds\le CN^{-1}$, shown in [GM10, Theorem 1.3], it follows that

$$
\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\|Z_s-Z_{t_i}\|_2^2 ds \le CN^{-1}+C\|\Phi\|_\infty N^{-2}\varepsilon^{3\theta_L/4-1}+CN^{-2\gamma/\beta}+C\varepsilon^{2\gamma}\big(1+\ln(N)\vee 1\big).
$$

Let $\delta\in[0,1]$ and set $\varepsilon=N^{-(1+\delta)/(2\gamma)}$ sufficiently large so that $N^{-\delta}\ln(N)\le 1$; for $N\in\{1,2\}$, it is sufficient to take $\delta=0$, and for $N>2$, $\delta=\ln\ln(N)/\ln(N)$. Recalling further that $2\gamma<\beta$, this implies that, under $(\mathbf{A_{b\Phi}})$ and $(\mathbf{A_{\partial f}})$,

$$
\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\|Z_s-Z_{t_i}\|_2^2 ds \le CN^{-1}+C\|\Phi\|_\infty N^{-2+(3\theta_L/4-1)(1+\delta)/(2\gamma)}.
$$

Recall the BSDE (Y_M,Z_M) from Corollary 1.4.9. The terminal condition and driver of (Y_M,Z_M) satisfy assumptions $(\mathbf{A_{b\Phi}})$ and $(\mathbf{A_{\partial f}})$, respectively, and it follows that

$$
\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\|Z_{M,s}-Z_{M,t_i}\|_2^2 ds \le CN^{-1}+CMN^{-2+(3\theta_L/4-1)(1+\delta)/(2\gamma)}.
$$

The proof is complete by taking $M=2\log(N)\vee 1$, R_t sufficiently large and applying Lemma 1.6.5. □

Remark 1.6.7. Thanks to Theorem 1.6.13, the $\mathbf{L}_2$-regularity of Z converges to zero with the optimal rate CN^{-1} if

$$8\gamma + 3\theta_L/(1+\delta_N) \geq 4 + 2\gamma \frac{\ln\ln(N)\mathbf{1}_{N\geq 3}}{\ln(N)}.$$

Since $\ln\ln(N)/\ln(N)$ is bounded above by 1 and δ_N is bounded above by 0.5, it suffices that $7\gamma + 2\theta_L \geq 4$. This is a different set of conditions to those imposed by Proposition 1.6.2. To take a specific example, let $\alpha = \theta_L$ and $\theta_c = 1$ (whence $\gamma = \alpha$), the case related to the quadratic BSDE (Section 1.3.1). It follows that $\gamma = \alpha$ and the optimal rate of convergence is obtained for $\alpha \geq 4/9$. This is a slight improvement on Proposition 1.6.2, which required that $\alpha \geq 1/2$, but the result requires more constraints on the integrability of $\Phi(X_T)$ and may depend on θ_c.

1.7 Discrete time approximation using Malliavin weights

In this section, we assume that, for each N, we are given Markov chains $(X_i)_{0\leq i\leq N}$, $(\nabla X_i)_{0\leq i\leq N}$ and $(\nabla X_i^{(-1)})_{0\leq i\leq N}$ on the time-grid $\pi_N^{(\beta)}$ such that

$$\max_i \|X_{t_i} - X_i\|_4 + \max_i \|\nabla X_{t_i} - \nabla X_i\|_4 + \max_i \ \|\nabla X_{t_i}^{(-1)} - \nabla X_i^{(-1)}\|_4 \leq C_X N^{-1/2}, \quad (1.7.1)$$

$$\max_i \|X_i\|_4 + \max_i \|\nabla X_{t_i} - \nabla X_i\|_4 + \max_i \ \|\nabla X_i^{(-1)}\|_4 \leq C_X, \quad (1.7.2)$$

$$\|\Phi(X_T) - \Phi(X_N)\|_2^2 \leq C_X N^{-1}, \quad (1.7.3)$$

for some constant C_X independent of N. From now on, a constant C is additionally allowed to depend on C_X. We will also need the driver f to be $\frac{1}{2}$-Hölder continuous in its time parameter: for all $(x, y, z) \in \mathbb{R}^d \times \mathbb{R} \times (\mathbb{R}^q)^\top$ and $(t, t') \in [0, T]^2$,

$$|f(t, x, y, z) - f(t', x, y, z)| \leq L_f |t - t'|^{1/2} \quad (1.7.4)$$

We use the following discrete-time approximation - built with the Markov chain approximations of X, ∇X and $\nabla X^{(-1)}$ - of the Malliavin weights (1.4.1):

$$H_j^i := \frac{1}{t_j - t_i}\Big(\sum_{k=i}^{j-1} (\sigma^{-1}(t_k, X_k)\nabla X_k \nabla X_i^{-1} \sigma(t_i, X_i))^\top \Delta W_k\Big)^\top \quad (1.7.5)$$

We propose the following discrete approximation $(Y_i, Z_i)_{0,\leq i\leq N-1}$ of (Y, Z), the BSDE (1.1.1), on the time-grid $\pi_N^{(\beta)}$: for $i = 0, \ldots, N-1$, set $f_i(x, y, z) := f(t_i, x, y, z)$, and define recursively $Y_N = \Phi(X_N)$, $Z_N = 0$, and

$$\left.\begin{aligned} Y_i &:= \mathbb{E}_i[\Phi(X_N) + \textstyle\sum_{j=i+1}^{N-1} f_j(X_j, Y_j, Z_j)\Delta_j], \\ Z_i &:= \mathbb{E}_i[\Phi(X_N)H_N^i + \textstyle\sum_{j=i+1}^{N-1} f_j(X_j, Y_j, Z_j)H_j^i\Delta_j]. \end{aligned}\right\} \quad (1.7.6)$$

where $\mathbb{E}_i[\cdot] := \mathbb{E}_{t_i}[\cdot]$. In the remainder of this section, we will estimate the error

$$\mathcal{E}^M(N) := \max_{0\leq i\leq N-1} \mathbb{E}[|Y_{t_i} - Y_i|^2] + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathbb{E}[|Z_t - Z_i|^2]dt. \quad (1.7.7)$$

associated to the Malliavin weights scheme (1.7.6). Following on from the representation of Theorem 1.5.1, we work with the version of Z such that

$$Z_t = \mathbb{E}_t[\Phi(X_T)H_T^t + \int_t^T f(s, X_s, Y_s, Z_s)H_s^t ds] \quad \text{for all } t \in [0, T) \text{ almost surely.}$$

It is in fact be sufficient to bound

$$\mathcal{E}^D(N) := \max_{0 \le i \le N-1} \mathbb{E}[|Y_{t_i} - Y_i|^2] + \sum_{i=0}^{N-1} \mathbb{E}[|Z_{t_i} - Z_i|^2]\Delta_i. \tag{1.7.8}$$

because the Young's inequality yields

$$\mathcal{E}^M(N) \le 2 \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathbb{E}[|Z_t - Z_{t_i}|^2]dt + 2\mathcal{E}^D(N) \tag{1.7.9}$$

and we may use the results of Proposition 1.6.2 or Theorem 1.6.13 to bound the first term.

Before we start with the estimation of $\mathcal{E}^D(N)$, we require some preliminary results.

Lemma 1.7.1. *There is a constant C such that, for all N and $t \in [t_i, t_{i+1}]$,*

$$\|\sigma^{-1}(t_i, X_{t_i}) - \sigma^{-1}(t_i, X_i)\|_4 + \|\sigma^{-1}(t, X_t) - \sigma^{-1}(t_i, X_{t_i})\|_4 \le CN^{-1/2}.$$

Proof. Since $\sigma^{-1}(t, \cdot)$ is Lipschitz continuous uniformly in t and $\sigma^{-1}(\cdot, x)$ is 1/2-Hölder continuous uniformly in x, from Lemma 1.1.4, the result follows from (1.7.1) combined with the usual bound $\|X_t - X_{t_i}\|_4 \le C\sqrt{t - t_i}$ for all $t \in [t_i, t_{i+1}]$. □

The next Proposition is based on Lemma 1.7.1 and is important for the discretization estimates later.

Proposition 1.7.2. *Let $(\mathbf{A_{b\Phi}})$ be in force. There is a constant C such that, for all N and $i \in \{0, \dots, N-1\}$,*

$$\|H_{t_j}^{t_i} - H_j^i\|_2 \le \frac{CN^{-1}}{\sqrt{t_j - t_i}}, \qquad \|H_j^i\|_2 \le \frac{C}{\sqrt{t_j - t_i}}, \tag{1.7.10}$$

$$\|\mathbb{E}_i[\Phi(X_T)(H_T^{t_i} - H_N^i)]\|_2 \le \frac{C\|\Phi\|_\infty N^{-1}}{\sqrt{T - t_i}} \tag{1.7.11}$$

$$\|\sum_{j=i+1}^{N-1} \mathbb{E}_i[f_j(X_{t_j}, Y_{t_j}, Z_{t_j})(H_{t_j}^{t_i} - H_j^i)]\Delta_j\|_2 \le CN^{-1/2}\|\Phi\|_\infty \big(1 \vee (T - t_i)^{(\theta_L \wedge (2\gamma) - 1)/2}\big) \tag{1.7.12}$$

Proof. In what follows, C may change from line to line.
For any $j > i$ and $t > t_i$, define

$$N_t^{t_i} := \sigma^{-1}(t, X_t)\nabla X_t \nabla X_{t_i}^{(-1)}\sigma(t_i, X_{t_i}) \quad \text{and} \quad N_j^i := \sigma^{-1}(t_j, X_j)\nabla X_j \nabla X_i^{(-1)}\sigma(t_i, X_i).$$

Observe the decompositions into telescopic sums

$$
\begin{aligned}
N_t^{t_i} - N_{t_j}^{t_i} &= \sigma^{-1}(t, X_t)(\nabla X_t - \nabla X_{t_j})\nabla X_{t_i}^{(-1)}\sigma(t_i, X_{t_i}) \\
&\quad + (\sigma^{-1}(t, X_t) - \sigma^{-1}(t_j, X_{t_j}))\nabla X_{t_j}\nabla X_{t_i}^{(-1)}\sigma(t_i, X_{t_i}), \\
N_{t_j}^{t_i} - N_j^i &= \sigma^{-1}(t_j, X_{t_j})\nabla X_{t_j}\nabla X_{t_i}^{(-1)}(\sigma(t_i, X_{t_i}) - \sigma(t_i, X_i)) \\
&\quad + \sigma^{-1}(t_j, X_{t_j})\nabla X_{t_j}(\nabla X_{t_i}^{(-1)}) - \nabla X_i^{(-1)})\sigma(t_i, X_i) \\
&\quad + \sigma^{-1}(t_j, X_{t_j})(\nabla X_{t_j} - \nabla X_j)\nabla X_{t_i}^{(-1)}\sigma(t_i, X_i) \\
&\quad + (\sigma^{-1}(t_j, X_{t_j}) - \sigma^{-1}(t_j, X_j))\nabla X_j \nabla X_{t_i}^{(-1)}\sigma(t_i, X_i)
\end{aligned}
$$

it follows from Lemma 1.7.1, the boundedness and Lipschitz continuity of σ and σ^{-1} (Lemma 1.1.4), the bounds (1.7.1) and (1.7.2), and the Cauchy-Schwarz inequality that for any $j > i$ and $t \in [t_j, t_{j+1}]$,

$$
\|N_t^{t_i} - N_{t_j}^{t_i}\|_2^2 + \|N_{t_j}^{t_i} - N_j^i\|_2^2 \leq CN^{-1}. \tag{1.7.13}
$$

Define $\hat{H}_j^i := \frac{1}{t_j - t_i}\big(\sum_{k=i}^{j-1}(N_{t_k}^{t_i})^\top \Delta W_k\big)^\top$. To show (1.7.10), apply Itô's isometry, the inequality (1.7.13), and Lemma 1.4.1:

$$
\begin{aligned}
\|H_{t_j}^{t_i} - H_j^i\|_2^2 &\leq 2\|H_{t_j}^{t_i} - \hat{H}_j^i\|_2^2 + 2\|H_j^i - \hat{H}_j^i\|_2^2 \\
&= 2\frac{\sum_{k=i}^{j-1}\int_{t_j}^{t_{j+1}} \|N_t^{t_i} - N_{t_k}^{t_i}\|_2^2 dt}{(t_j - t_i)^2} + 2\frac{\sum_{k=i}^{j-1}\|N_{t_k}^{t_i} - N_k^i\|_2^2 \Delta_k}{(t_j - t_i)^2} \leq \frac{CN^{-1}}{t_j - t_i}, \\
\|H_j^i\|_2 &\leq \|H_{t_j}^{t_i} - H_j^i\|_2 + \|H_{t_j}^{t_i}\|_2 \leq \frac{C}{\sqrt{t_j - t_i}}.
\end{aligned}
$$

The estimate (1.7.11) is obtained from the estimate (1.7.10) and the bound on $\Phi(x)$:

$$
\|\mathbb{E}_i[\Phi(X_T)(H_T^{t_i} - H_N^i)]\|_2 \leq \|\Phi\|_\infty \|H_T^{t_i} - H_N^i\|_2 \leq C\|\Phi\|_\infty N^{-1}(T - t_i)^{-1/2}.
$$

For (1.7.12), use Lemma 1.4.5 with $(Y_1, Z_1) = (0, 0)$ and $(Y_2, Z_2) = (Y, Z)$ combined with the bounds on $\Phi(x)$ and the local bound on f of (1.1.3) to obtain the bounds

$$
|Y_t| \leq C\|\Phi\|_\infty + C\int_t^T \frac{dr}{(T - r)^{1-\theta_c}}, \qquad |Z_t| \leq \frac{C\|\Phi\|_\infty}{\sqrt{T - t}} + C\int_t^T \frac{dr}{(T - r)^{1-\theta_c}\sqrt{r - t}} \quad \text{for all } t \in [0, T) \text{ } a.s.
$$

This implies, from the local Lipschitz continuity and boundedness of f in (1.1.3) that

$$
|f_j(X_{t_j}, Y_{t_j}, Z_{t_j})| \leq \frac{C\|\Phi\|_\infty}{(T - t_j)^{1-\theta_L/2}} + \frac{C\int_{t_j}^T (T - r)^{\theta_c - 1}(r - t_j)^{-1/2} dr}{(T - t_j)^{(1-\theta_L)/2}} + \frac{C}{(T - t_j)^{1-\theta_c}} \leq \frac{C\|\Phi\|_\infty}{(T - t_j)^{1-(2\theta_c)\wedge\theta_L/2}}
$$

where the integral $\int_t^T (T - r)^{\theta_c - 1}(r - t_j)^{-1/2} dr$ is bounded above by $C(T - t_j)^{(2\theta_c - 1)/2}$ using Lemma 1.8.5. Combining this with the estimate (1.7.10) and the local Lipschitz continuity of f in (1.1.3)

$$
\|\sum_{j=i+1}^{N-1} \mathbb{E}_i[f_j(X_{t_j} Y_{t_j}, Z_{t_j})(H_{t_j}^{t_i} - H_j^i)]\Delta_j\|_2 \leq \sum_{j=i+1}^{N-1} \frac{C\|\Phi\|_\infty N^{-1/2}\Delta_j}{(T - t_j)^{1-(2\gamma)\wedge\theta_L/2}\sqrt{t_j - t_i}}
$$

and apply Lemma 1.8.5. □

Lemma 1.7.3. *For all $t_i, t_j \in \pi$ such that $t_i \leq t_j$ and $r \in [t_j, T]$,*

$$\mathbb{E}_i[f(r, X_r, Y_r, Z_r)H_r^{t_i}] = \mathbb{E}_i[f(r, X_r, Y_r, Z_r)H_{t_j}^{t_i}]. \tag{1.7.14}$$

Moreover,

$$\begin{aligned}\mathbb{E}_i[\int_{t_i}^{T} f(r, X_r, Y_r, Z_r)H_r^{t_i} dr] &= \mathbb{E}_i[\sum_{j=i+1}^{N-1} f_j(X_{t_j}, Y_{t_j}, Z_{t_j})H_{t_j}^{t_i}\Delta_j] + \mathbb{E}_i[\int_{t_i}^{t_{i+1}} f(r, X_r, Y_r, Z_r)H_r^{t_i} dr] \\ &\quad + \mathbb{E}_i[\sum_{j=i+1}^{N-1}\int_{t_j}^{t_{j+1}} (f(r, X_r, Y_r, Z_r) - f_j(X_{t_j}, Y_{t_j}, Z_{t_j}))H_{t_j}^{t_i} dr].\end{aligned} \tag{1.7.15}$$

Proof. The arguments are the same as for the proof of Theorem 1.5.1, except that in (1.5.2) the integral with respect to dt is over $v \in [t_i, t_j]$:

$$\begin{aligned}&\frac{1}{t_j - t_i}\int_{t_i}^{t_j} D_v f^{(\varepsilon)}(\Theta_r)\sigma^{-1}(v, X_v)\nabla X_v dv \\ &\quad = f_x^{(\varepsilon)}(\Theta_r)\nabla X_r + f_y^{(\varepsilon)}(\Theta_r)(u(r, X_r)\nabla X_r + \nabla y_r^{(\varepsilon)}) \\ &\qquad + f_z^{(\varepsilon)}(\Theta_r)(U(r, X_r)\nabla X_r + \nabla z_r^{(\varepsilon)}) \qquad\qquad m \times \mathbb{P} - a.e.\end{aligned}$$

The relation (1.7.15) is now straightforward to obtain. □

Lemma 1.7.4. *There is a constant $C \geq 0$ such that, for all N and $i \in \{0, \ldots, N-1\}$,*

$$\|\mathbb{E}_{t_i}[\int_{t_i}^{t_{i+1}} f(r, X_r, Y_r, Z_r)H_r^{t_i} dr]\|_2 \leq \int_{t_i}^{t_{i+1}} \frac{C dr}{(T-r)^{1-\gamma}\sqrt{r - t_i}}, \tag{1.7.16}$$

$$\begin{aligned}&\|\sum_{j=i+1}^{N-1} \mathbb{E}_{t_i}[(f(t_j, X_{t_j}, Y_{t_j}, Z_{t_j}) - f_j(X_j, Y_j, Z_j))H_{t_j}^{t_i}]\Delta_{j-1}\|_2 \\ &\qquad \leq CN^{-1/2} + C\sum_{j=i+1}^{N-1} \frac{\|Y_{t_j} - Y_j\|_2 + \|Z_{t_j} - Z_j\|_2}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}}\Delta_j,\end{aligned} \tag{1.7.17}$$

$$\begin{aligned}&\|\sum_{j=i+1}^{N-1} \mathbb{E}_{t_i}[\int_{t_j}^{t_{j+1}} (f(r, X_r, Y_r, Z_r) - f(t_j, X_{t_j}, Y_{t_j}, Z_{t_j}))H_{t_j}^{t_i} dr]\|_2 \\ &\qquad \leq CN^{-1/2} + C\sum_{j=i+1}^{N-1} \frac{\int_{t_j}^{t_{j+1}}\{\|Y_r - Y_{t_j}\|_2 + \|Z_r - Z_{t_j}\|_2\}dr}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}}.\end{aligned} \tag{1.7.18}$$

Proof. In what follows, C may change from line to line.
Using Lemma 1.4.1 and the moment bounds $\|Y_t\|_2 \leq C$ and $\|Z_t\|_2 \leq C\big(1 \vee (T-t)^{(2\theta_c \wedge \alpha - 1)/2}\big)$

of Corollary 1.4.4,

$$
\begin{aligned}
&\|\mathbb{E}_{t_i}[\int_{t_i}^{t_{i+1}} f(r, X_r, Y_r, Z_r)H_r^{t_i}]dr\|_2 \\
&\le \int_{t_i}^{t_{i+1}} \big(\|f(r, X_r, Y_r, Z_r) - f(r, X_r, 0, 0)\|_2 + \|f(r, X_r, 0, 0)\|_2\big)\|H_r^{t_i}\|_2 dr \\
&\le C\int_{t_i}^{t_{i+1}} \frac{dr}{(T-r)^{1-\gamma}\sqrt{r-t_i}}
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&\|\sum_{j=i+1}^{N-1} \mathbb{E}_{t_i}[\big(f_j(X_{t_j}, Y_{t_j}, Z_{t_j}) - f_j(X_j, Y_j, Z_j)\big)H_{t_j}^{t_i}]\Delta_{j-1}\|_2 \\
&\le C\sum_{j=i+1}^{N-1} \frac{\|X_{t_j} - X_j\|_2 + \|Y_{t_j} - Y_j\|_2 + \|Z_{t_j} - Z_j\|_2}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}}\Delta_{j-1} \\
&\le C\sum_{j=i+1}^{N-1} \frac{\|Y_{t_j} - Y_j\|_2 + \|Z_{t_j} - Z_j\|_2}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}}\Delta_j
\end{aligned}
$$

For (1.7.18), the t-Hölder continuity of f in (1.7.4), the Cauchy-Schwarz inequality, Minkowski's inequality, and Hölder's inequality are needed:

$$
\begin{aligned}
&\|\sum_{j=i+1}^{N-1} \mathbb{E}_{t_i}[\int_{t_j}^{t_{j+1}} \big(f(r, X_r, Y_r, Z_r) - f_j(X_{t_j}, Y_{t_j}, Z_{t_j})\big)H_{t_j}^{t_i} dr]\|_2 \\
&\le C\sum_{j=i+1}^{N-1} \frac{\int_{t_j}^{t_{j+1}} \|f(r, X_r, Y_r, Z_r) - f_j(X_r, Y_r, Z_r)\|_2 dr}{\sqrt{t_j - t_i}} \\
&\quad + C\sum_{j=i+1}^{N-1} \frac{\int_{t_j}^{t_{j+1}} \|f_j(X_r, Y_r, Z_r) - f_j(X_{t_j} Y_{t_j}, Z_{t_j})\|_2 dr}{\sqrt{t_j - t_i}} \\
&\le C\sum_{j=i+1}^{N-1} \frac{\int_{t_j}^{t_{j+1}} \sqrt{r - t_j} dr}{\sqrt{t_j - t_i}} + C\sum_{j=i+1}^{N-1} \frac{\int_{t_j}^{t_{j+1}} \{\|X_r - X_{t_j}\|_2 + \|Y_r - Y_{t_j}\|_2 + \|Z_r - Z_{t_j}\|_2\} dr}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}}
\end{aligned}
$$

□

We see in (1.7.18) of Lemma 1.7.4 the appearance of the terms

$$
\sum_{j=i+1}^{N-1} \frac{\int_{t_j}^{t_{j+1}} \{\|Y_r - Y_{t_j}\|_2 + \|Z_r - Z_{t_j}\|_2\} dr}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}}
$$

In the following proposition, we obtain a bound that is intrinsically related to approximation error caused by these terms. Proposition 1.6.3 will be essential in the proof of this result.

Proposition 1.7.5. *Let* $(\mathbf{A}_{\mathbf{exp}\boldsymbol{\Phi}})$ *be in force and suppose that* $0 < \beta < (2\gamma) \wedge \alpha \wedge \theta_L$. *There is a* *constant C such that, for all N,*

$$\Big(\sum_{j=0}^{N-1}\frac{\int_{t_j}^{t_{j+1}}\{\|Y_r-Y_{t_j}\|_2+\|Z_r-Z_{t_j}\|_2\}dr}{(T-t_j)^{(1-\theta_L)/2}}\Big)^2+\sum_{i=0}^{N-2}\Big(\sum_{j=i+1}^{N-1}\frac{\int_{t_j}^{t_{j+1}}\{\|Y_r-Y_{t_j}\|_2+\|Z_r-Z_{t_j}\|_2\}dr}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j-t_i}}\Big)^2\Delta_i$$

$$\leq CN^{-1}+CN^{(1-3\theta_L/4)/(2\gamma)-2}(\ln(N)\vee 1). \qquad (1.7.19)$$

Proof. We will prove the bounds for

$$\sum_{i=0}^{N-2}\Big(\sum_{j=i+1}^{N-1}\frac{\int_{t_j}^{t_{j+1}}\{\|Y_r-Y_{t_j}\|_2+\|Z_r-Z_{t_j}\|_2\}dr}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j-t_i}}\Big)^2\Delta_i$$

The bounds for

$$\Big(\sum_{j=0}^{N-1}\frac{\int_{t_j}^{t_{j+1}}\{\|Y_r-Y_{t_j}\|_2+\|Z_r-Z_{t_j}\|_2\}dr}{(T-t_j)^{(1-\theta_L)/2}}\Big)^2$$

are obtained analogously. Moreover, we will only prove the result for the terms in Z. The bound for the terms in Y are also obtained analogously.

In what follows, C may change from line to line.

We first prove the result under ($\mathbf{A_{\partial f}}$) and ($\mathbf{A_{b\Phi}}$), and then obtain the general result by means of mollification. Recall the BSDE $(Y^{(\varepsilon)},Z^{(\varepsilon)})$ from Definition 1.2.9 in Section 1.2.2. The triangle inequality yields $\|Z_t-Z_{t_i}\|_2\leq\|Z_t-Z_t^{(\varepsilon)}\|_2+\|Z_{t_i}-Z_{t_i}^{(\varepsilon)}\|_2+\|Z_t^{(\varepsilon)}-Z_{t_i}^{(\varepsilon)}\|_2$. This implies that

$$\sum_{j=i+1}^{N-1}\frac{\int_{t_j}^{t_{j+1}}\|Z_r-Z_{t_j}\|_2dr}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j-t_i}}\leq\sum_{j=i+1}^{N-1}\frac{\int_{t_j}^{t_{j+1}}\{\|Z_t-Z_t^{(\varepsilon)}\|_2+\|Z_{t_i}-Z_{t_i}^{(\varepsilon)}\|_2+\|Z_t^{(\varepsilon)}-Z_{t_i}^{(\varepsilon)}\|_2\}dr}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j-t_i}}.$$

To bound the terms in $Z-Z^{(\varepsilon)}$, recall the bound (1.4.14) from Corollary 1.4.8 gives

$$\|Z_t-Z_t^{(\varepsilon)}\|_2\leq C\int_{t\vee(T-\varepsilon)}^T\frac{dr}{(T-r)^{1-\gamma}\sqrt{r-t}}\qquad\forall\ t\in[0,T).$$

For $j\leq N-2$, the bound on $\|Z_t-Z_t^{(\varepsilon)}\|_2$ implies that

$$\int_{t_j}^{t_{j+1}}\|Z_t-Z_t^{(\varepsilon)}\|_2dt\leq C\int_{t_j}^{t_{j+1}}\frac{\int_{T-\varepsilon}^T(T-r)^{\gamma-1}dr}{\sqrt{t_{N-1}-t}}dt\leq C\varepsilon^\gamma\Delta_j\leq C\varepsilon^\gamma\int_{t_j}^{t_{j+1}}\frac{dt}{\sqrt{t_{N-1}-t}}.$$

Direct computation of the integral term yields

$$\begin{aligned}\int_{t_j}^{t_{j+1}}\frac{dt}{\sqrt{t_{N-1}-t}}&=2\big\{(t_{N-1}-t_j)^{1/2}-(t_{N-1}-t_{j+1})^{1/2}\big\}\\&\leq 2\Big\{\frac{t_{N-1}-t_j}{(t_{N-1}-t_j)^{1/2}}-\frac{t_{N-1}-t_{j+1}}{(t_{N-1}-t_j)^{1/2}}\Big\}=\frac{2\Delta_j}{(t_{N-1}-t_j)^{1/2}}\end{aligned}$$

Combining the above bounds and applying Lemma 1.8.5 implies that

$$\sum_{j=i+1}^{N-2} \frac{\int_{t_j}^{t_{j+1}} \|Z_t - Z_t^{(\varepsilon)}\|_2 dt}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}} \leq C\varepsilon^\gamma \sum_{j=i+1}^{N-2} \frac{\Delta_j}{(t_{N-1}-t_j)^{1-\theta_L/2}\sqrt{t_j - t_i}} \leq \frac{C\varepsilon^\gamma}{(t_{N-1}-t_i)^{(1-\theta_L)/2}}. \tag{1.7.20}$$

For the outstanding term, $j = N-1$, we implement Lemma 1.8.5 to show that

$$\int_{t_{N-1}}^T \|Z_t - Z_t^{(\varepsilon)}\|_2 dt \leq C \int_{t_{N-1}}^T \Big\{\int_t^T (T-r)^{\gamma-1}(r-t)^{-1/2} dr\Big\} dt \leq C\Delta_{N-1}^{1/2+\gamma},$$

whence it follows that

$$\frac{\int_{t_{N-1}}^T \|Z_t - Z_t^{(\varepsilon)}\|_2 dt}{\Delta_{N-1}^{(1-\theta_L)/2}\sqrt{t_{N-1}-t_i}} \leq \frac{\Delta_{N-1}^{\gamma+\theta_L/2}}{\sqrt{t_{N-1}-t_i}} \tag{1.7.21}$$

Combining (1.7.20) and (1.7.21), it follows that

$$\begin{aligned}
&\sum_{i=0}^{N-2}\bigg(\sum_{j=i+1}^{N-1} \frac{\int_{t_j}^{t_{j+1}} \|Z_t - Z_t^{(\varepsilon)}\|_2 dr}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}}\bigg)^2 \Delta_i \\
&\leq C\varepsilon^{2\gamma} \sum_{i=0}^{N-2} \frac{\Delta_i}{(t_{N-1}-t_i)^{1-\theta_L}} + C\Delta_{N-1}^{2\gamma+\theta_L} \sum_{i=0}^{N-2} \frac{\Delta_i}{t_{N-1}-t_i} \leq C\varepsilon^{2\gamma} + CN^{-2}\big(1+\ln(N)\big)
\end{aligned} \tag{1.7.22}$$

where we have used that $\Delta_{N-1}^{2\gamma+\theta_L} = TN^{(2\gamma+\theta_L)/\beta}$ and $\beta < (2\gamma)\wedge\theta_L$. Analogously, we can also show that

$$\sum_{i=0}^{N-2}\bigg(\sum_{j=i+1}^{N-1} \frac{\|Z_{t_j} - Z_{t_j}^{(\varepsilon)}\|_2 \Delta_j}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}}\bigg)^2 \Delta_i \leq C\varepsilon^{2\gamma} + CN^{-2}\big(1+\ln(N)\big). \tag{1.7.23}$$

Recalling the BSDEs (y,z) and $(y^{(\varepsilon)}, z^{(\varepsilon)})$ from Definition 1.2.9 and that $Z^{(\varepsilon)} = z + z^{(\varepsilon)}$, the triangle inequality yields $\|Z_t^{(\varepsilon)} - Z_{t_i}^{(\varepsilon)}\|_2 \leq \|z_t - z_{t_i}\|_2 + \|z_t^{(\varepsilon)} - z_{t_i}^{(\varepsilon)}\|_2$. In the proof of [GM10, Theorem 1.1], it is shown that, for all $t \in [0,T]$,

$$\|z_t - z_{t_i}\|_2^2 \leq \frac{C(t-t_i)}{(T-t)^{1-\alpha}} + C\int_{t_i}^t \frac{dr}{(T-r)^{2-\alpha}}.$$

Now, applying Jensen's inequality, Lemma 1.8.5, and the above bound, one obtains

$$\begin{aligned}
&\sum_{i=0}^{N-2}\bigg(\sum_{j=i+1}^{N-1} \frac{\int_{t_j}^{t_{j+1}} \|z_r - z_{t_j}\|_2 dr}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}}\bigg)^2 \Delta_i \\
&\leq \sum_{i=0}^{N-2}\bigg(\sum_{j=i+1}^{N-1} \frac{C\Delta_j \int_{t_j}^{t_{j+1}} (t_{j+1}-r)^{(\alpha-1)/2} dr + C\int_{t_j}^{t_{j+1}} \Big(\int_{t_j}^r (T-t)^{\alpha-2} dt\Big)^{1/2} dr}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}}\bigg)^2 \Delta_i \\
&\leq \sum_{i=0}^{N-2}\bigg(\sum_{j=i+1}^{N-1} \frac{C\Delta_j^{(3+\alpha)/2} + C\Delta_j^{1/2}\Big(\int_{t_j}^{t_{j+1}} (t_{j+1}-t)(T-t)^{\alpha-2} dt\Big)^{1/2}}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}}\bigg)^2 \Delta_i.
\end{aligned} \tag{1.7.24}$$

For $j \leq N-2$, one can apply Lemma 1.8.4 to show that

$$\int_{t_j}^{t_{j+1}} (t_{j+1}-t)(T-t)^{\alpha-2}dt \leq \frac{\Delta_j}{(T-t_{j+1})^{1-\beta}} \int_{t_j}^{t_{j+1}} (T-t)^{\alpha-\beta-1}dt \leq \frac{CN^{-1}\Delta_j}{(T-t_j)^{1-\alpha+\beta}}.$$

where we have used Lemma 1.8.4 to show that $(\Delta_j/\Delta_{j+1})\Delta_{j+1}(T-t_{j+1})^{\beta-1} \leq CN^{-1}$ and the direct computation

$$\begin{aligned}\int_{t_j}^{t_{j+1}} \frac{dt}{(T-t)^{1-\alpha+\beta}} &= \frac{1}{\alpha-\beta}\big\{(T-t_j)^{\alpha-\beta} - (T-t_{j+1})^{\alpha-\beta}\big\} \\ &\leq \frac{1}{\alpha-\beta}\Big\{\frac{T-t_j}{(T-t_j)^{1-\alpha+\beta}} - \frac{T-t_{j+1}}{(T-t_j)^{1-\alpha+\beta}}\Big\} = \frac{\Delta_j}{(\alpha-\beta)(T-t_j)^{1-\alpha+\beta}}\end{aligned}$$

On the other hand, for $j = N-1$, since $\beta < \alpha$,

$$\int_{t_{N-1}}^{T} (T-t)(T-t)^{\alpha-2}dt = \Delta_{N-1}^{\alpha} \leq TN^{-1}.$$

Substituting these bounds into (1.7.24) and implementing Lemma 1.8.5, we obtain

$$\begin{aligned}&\sum_{i=0}^{N-2}\Bigg(\sum_{j=i+1}^{N-1}\frac{\int_{t_j}^{t_{j+1}}\|z_r - z_{t_j}\|_2 dr}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j-t_i}}\Bigg)^2\Delta_i \\ &\leq CN^{-1} + CN^{-1}\sum_{i=0}^{N-2}\Bigg(\sum_{j=i+1}^{N-2}\frac{\Delta_j}{(T-t_j)^{1+(\beta-\alpha-\theta_L)/2}\sqrt{t_j-t_i}}\Bigg)^2\Delta_i + CN^{-1}\sum_{i=0}^{N-2}\Bigg(\frac{\Delta_{N-1}^{\theta_L/2}}{\sqrt{t_{N-1}-t_i}}\Bigg)^2\Delta_i \\ &\leq CN^{-1} + CN^{-1}\sum_{i=0}^{N-2}\frac{\Delta_i}{(t_{N-1}-t_i)^{1-\theta_L}} \leq CN^{-1}. \qquad (1.7.25)\end{aligned}$$

In the bounds (1.6.3), we used the inequality

$$\|z_r^{(\varepsilon)} - z_{t_i}^{(\varepsilon)}\|_2 \leq C\int_{t_i}^{r}\|a_t^{(\varepsilon)}\|_2 dt + C\int_{t_i}^{r}\|V_t^{(\varepsilon)}\|_2 dt + C\Delta_i^{1/2}. \qquad (1.7.26)$$

Using $\|a_t^{(\varepsilon)}\|_2 \leq C(T-t)^{(\alpha+\theta_L-3)/2}$, as shown Lemma 1.2.14, it follows that

$$\begin{aligned}&\sum_{j=i+1}^{N-1}\frac{\int_{t_j}^{t_{j+1}}\{\int_{t_j}^{r}\|a_t^{(\varepsilon)}\|_2 dt\}dr}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j-t_i}} \leq C\sum_{j=i+1}^{N-1}\frac{\int_{t_j}^{t_{j+1}}\{\int_{t_j}^{r}(T-t)^{(\theta_L+\alpha-3)/2}dt\}dr}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j-t_i}} \\ &\leq C\sum_{j=i+1}^{N-2}\frac{\int_{t_j}^{t_{j+1}}(T-r)^{(\alpha-2)/2}dr\int_{t_i}^{t_{j+1}}(T-t)^{(\theta_L-1)/2}dt}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j-t_i}} + C\frac{\int_{t_{N-1}}^{T}(T-r)^{(\theta_L+\alpha-1)/2}dr}{(T-t_{N-1})^{(1-\theta_L)/2}\sqrt{t_{N-1}-t_i}}\end{aligned}$$

where we have used

$$\int_{t_{N-1}}^{T}\{\int_{t_{N-1}}^{r}(T-t)^{(\alpha+\theta_L-3)/2}dt\}dr = \frac{\int_{t_{N-1}}^{T}(T-t)^{(\alpha+\theta_L-1)/2}dt}{(1-\alpha-\theta_L)/2}.$$

As before, we use direct computation to show

$$\int_{t_j}^{t_{j+1}} \frac{dt}{(T-t)^{1-\alpha/2}} = \frac{2}{\alpha}\big\{(T-t_j)^{\alpha/2} - (T-t_{j+1})^{\alpha/2}\big\}$$
$$\leq \frac{2}{\alpha}\Big\{\frac{T-t_j}{(T-t_j)^{1-\alpha/2}} - \frac{T-t_{j+1}}{(T-t_j)^{1-\alpha/2}}\Big\} = \frac{2\Delta_j}{\alpha(T-t_j)^{1-\alpha/2}}$$

whence it follows that

$$\sum_{j=i+1}^{N-1} \frac{\int_{t_j}^{t_{j+1}}\{\int_{t_j}^{r}\|a_t^{(\varepsilon)}\|_2 dt\}dr}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j-t_i}} \leq C\Big(\max_{0\leq i\leq N-1}\frac{\Delta_i^{1+\theta_L}}{(T-t_i)^{1-\beta}}\Big)^{1/2}\sum_{j=i+1}^{N-2}\frac{\Delta_j}{(T-t_j)^{(2+\beta-\theta_L-\alpha)/2}\sqrt{t_j-t_i}}$$
$$+ C\frac{\Delta_{N-1}^{(1+\alpha+\theta_L)/2}}{(T-t_{N-1})^{(1-\theta_L)/2}\sqrt{t_{N-1}-t_i}}$$
$$\leq \frac{CN^{-1/2}}{(T-t_i)^{(1+\beta-\theta_L-\alpha)/2}} + \frac{CN^{-(\alpha+2\theta_L)/(2\beta)}}{\sqrt{t_{N-1}-t_i}}. \tag{1.7.27}$$

On the other hand, we obtain from Proposition 1.6.3 that $\|V_t^{(\varepsilon)}\|_2 \leq \phi(t,\varepsilon,\theta_L)$. It follows from Lemma 1.8.5 that for all j and $r \in [t_j, t_{j+1}]$,

$$\int_{t_j}^{r}\|V_t^{(\varepsilon)}\|_2 dt \leq C\|\Phi\|_\infty \int_{t_j}^{r}\Big\{\int_t^{T-\varepsilon}(T-s)^{(\theta_L-3)/2}(s-t)^{-1/2}ds\Big\}dt$$
$$\leq C\|\Phi\|_\infty \varepsilon^{(3\theta_L/4-1)/2}\int_{t_j}^{r}\Big\{\int_t^{T}(T-s)^{\theta_L/8-1}(s-t)^{-1/2}ds\Big\}dt$$
$$\leq C\|\Phi\|_\infty \varepsilon^{(3\theta_L/4-1)/2}\int_{t_j}^{r}(T-t)^{(\theta_L/4-1)/2}dt \leq C\|\Phi\|_\infty\frac{\varepsilon^{(3\theta_L/4-1)/2}\Delta_j}{(T-t_j)^{(1-\theta_L/4)/2}}$$

where we have used the direct computations

$$\int_{t_j}^{r}\frac{dt}{(T-t)^{(1-\theta_L/4)/2}} = \frac{2}{1+\theta_L/4}\big\{(T-t_j)^{(1+\theta_L/4)/2} - (T-r)^{(1+\theta_L/4)/2}\big\}$$
$$\leq \frac{2}{1+\theta_L/4}\Big\{\frac{T-t_j}{(T-t_j)^{(1+\theta_L/4)/2}} - \frac{T-r}{(T-t_j)^{(1+\theta_L/4)/2}}\Big\} \leq \frac{C\Delta_j}{(T-t_j)^{(1+\theta_L/4)/2}}$$

Therefore, using Lemma 1.8.5 and the above bound,

$$\sum_{j=i+1}^{N-1}\frac{\int_{t_j}^{t_{j+1}}\{\int_{t_j}^{r}\|V_t^{(\varepsilon)}\|_2 dt\}dr}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j-t_i}} \leq C\|\Phi\|_\infty \varepsilon^{(3\theta_L/4-1)/2}\max_i \Delta_i \sum_{j=i+1}^{N-1}\frac{\Delta_j}{(T-t_j)^{1-5\theta_L/8}\sqrt{t_j-t_i}} \tag{1.7.28}$$
$$\leq \frac{C\|\Phi\|_\infty \varepsilon^{(3\theta_L/4-1)/2}N^{-1}}{(T-t_i)^{(1-5\theta_L/4)/2}}. \tag{1.7.29}$$

Now, substituting (1.7.27) and (1.7.29) into (1.7.26), it follows that

$$\sum_{i=0}^{N-2}\Big(\sum_{j=i+1}^{N-1}\frac{\int_{t_j}^{t_{j+1}}\|z_r^{(\varepsilon)}-z_{t_j}^{(\varepsilon)}\|_2dr}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j-t_i}}\Big)^2\Delta_i \le \sum_{i=0}^{N-2}\frac{CN^{-1}\Delta_i}{(T-t_i)^{1+\beta-\theta_L-\alpha}}+\sum_{i=0}^{N-2}\frac{CN^{-3}\Delta_i}{t_{N-1}-t_i}$$
$$+C\|\Phi\|_\infty\varepsilon^{3\theta_L/4-1}N^{-2}\sum_{i=0}^{N-2}\frac{\Delta_i}{(T-t_i)^{1-5\theta_L/4}}$$
$$\le CN^{-1}+CN^{-3}\big(1+\ln(N)\big)+C\|\Phi\|_\infty\varepsilon^{3\theta_L/4-1}N^{-2}. \tag{1.7.30}$$

Combining (1.7.22), (1.7.23), (1.7.25) and (1.7.30) yields

$$\sum_{i=0}^{N-2}\Big(\sum_{j=i+1}^{N-1}\frac{\int_{t_j}^{t_{j+1}}\|Z_r-Z_{t_j}\|_2dr}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j-t_i}}\Big)^2\Delta_i \le CN^{-1}+C\varepsilon^{2\gamma}+CN^{-2}\big(1+\ln(N)\big)+C\|\Phi\|_\infty\varepsilon^{3\theta_L/4-1}N^{-2}$$

and we take $\varepsilon = N^{-1/(2\gamma)}$ to complete the proof under $(\mathbf{A_{\partial f}})$ and $(\mathbf{A_{b\Phi}})$.

To prove the result without $(\mathbf{A_{\partial f}})$ or $(\mathbf{A_{b\Phi}})$, recall the mollified BSDE (Y_M, Z_M) from Corollary 1.4.9. Analogously to Theorem 1.6.6, the result follows from the triangle inequality and taking $M = C\ln(N)\vee 1$. □

We come to the main result of this section, namely the estimation of the error $\mathcal{E}^D(N)$ in (1.7.8).

Theorem 1.7.6. *Let* $(\mathbf{A_{exp\Phi}})$ *be and force and suppose that* $0<\beta\le 2\gamma\wedge\alpha\wedge\theta_L$*. There is a constant* C *such that, for all* N*,*

$$\mathcal{E}^D(N) := \max_{0\le i\le N-1}\|Y_{t_i}-Y_i\|_2^2+\sum_{i=0}^{N-1}\|Z_{t_i}-Z_i\|_2^2\Delta_i \le CN^{-1}+C\big(N^{-2\gamma}+N^{(1-3\theta_L/4)/(2\gamma)-2}\big)(\ln(N)\vee 1). \tag{1.7.31}$$

Proof In what follows, C may change from line to line. Fix $i\in\{0,\dots,N-1\}$. Using (1.7.15) from Lemma 1.7.3, it follows that

$$\|Z_{t_i}-Z_i\|_2 = \|\mathbb{E}_i[\Phi(X_T)H_T^{t_i}-\Phi(X_N)H_N^i+\int_{t_i}^T f(t,Y_t,Z_t)H_t^{t_i}dt-\sum_{j=i+1}^{N-1}f_j(Y_j,Z_j)H_j^i\Delta_j]\|_2$$
$$\le \|\mathbb{E}_i[\int_{t_i}^{t_{i+1}} f(r,Y_r,Z_r)H_r^{t_i}dr]\|_2+\|\mathbb{E}_i[\Phi(X_T)(H_T^{t_i}-H_N^i)]\|_2$$
$$+\frac{C\|\Phi(X_T)-\Phi(X_N)\|_2}{\sqrt{T-t_i}}+\|\sum_{j=i+1}^{N-1}\mathbb{E}_i[f(t_j,Y_{t_j},Z_{t_j})(H_{t_j}^{t_i}-H_j^i)]\Delta_j\|_2$$
$$+\|\sum_{j=i+1}^{N-1}\mathbb{E}_i[(f(t_j,Y_{t_j},Z_{t_j})-f_j(Y_j,Z_j))H_{t_j}^{t_i}]\Delta_j\|_2$$
$$+\|\sum_{j=i+1}^{N-1}\mathbb{E}_i[\int_{t_j}^{t_{j+1}}(f(r,Y_r,Z_r)-f(t_j,Y_{t_j},Z_{t_j}))H_{t_j}^{t_i}dr]\|_2 \tag{1.7.32}$$

Substituting (1.7.10) - (1.7.12) from Proposition 1.7.2, (1.7.16) - (1.7.18) from Lemma 1.7.4 into

(1.7.32), it follows that

$$\|Z_{t_i} - Z_i\|_2 \le \frac{C\|\Phi\|_\infty N^{-1/2}}{\sqrt{T-t_i}} + C\int_{t_i}^{t_{i+1}} \frac{dr}{(T-r)^{1-\gamma}\sqrt{r-t_i}} + \mathcal{E}(i)$$
$$+ C\sum_{j=i+1}^{N-1} \frac{\|Y_{t_j} - Y_j\|_2 + \|Z_{t_j} - Z_j\|_2}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}}\Delta_j \tag{1.7.33}$$

where

$$\mathcal{E}(i) := C\sum_{j=i+1}^{N-1} \frac{\Psi(j)}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j-t_i}}, \quad \Psi(j) := C\int_{t_j}^{t_{j+1}} \{\|Y_r - Y_{t_j}\|_2 + \|Z_r - Z_{t_j}\|_2\}dr.$$

Using a similar technique, $\|Y_{t_i} - Y_i\|_2$ is bounded above by

$$\|\Phi(X_T) - \Phi(X_N)\|_2 + C\int_{t_i}^{t_{i+1}} \frac{dr}{(T-r)^{1-\gamma}} + C\sum_{j=i+1}^{N-1} \frac{\Psi(j)}{(T-t_j)^{(1-\theta_L)/2}} + C\sum_{j=i+1}^{N-1} \frac{\|Y_{t_j} - Y_j\|_2 + \|Z_{t_j} - Z_j\|_2}{(T-t_j)^{(1-\theta_L)/2}}\Delta_j$$
$$\le \quad CN^{-1/2} + C\sum_{j=i+1}^{N-1} \frac{\Psi(j)}{(T-t_j)^{(1-\theta_L)/2}} + C\sum_{j=i+1}^{N-1} \frac{\|Y_{t_j} - Y_j\|_2 + \|Z_{t_j} - Z_j\|_2}{(T-t_j)^{(1-\theta_L)/2}}\Delta_j \tag{1.7.34}$$

where we have used the following direct computation to bound the first integral term:

$$\int_{t_i}^{t_{i+1}} \frac{dt}{(T-t)^{1-\gamma}} = \frac{1}{\gamma}\big\{(T-t_i)^\gamma - (T-t_{i+1})^\gamma\big\}$$
$$\le \frac{1}{\gamma}\Big\{\frac{T-t_i}{(T-t_i)^{1-\gamma}} - \frac{T-t_{i+1}}{(T-t_i)^{1-\gamma}}\Big\} = \frac{\Delta_j}{\gamma(T-t_i)^{1-\gamma}} \le CN^{-1/2}.$$

Define $\Theta_i := \|Y_{t_i} - Y_i\|_2 + \|Z_{t_i} - Z_i\|_2$. It follows from (1.7.33) and (1.7.34) that

$$\Theta_i \le \frac{C\|\Phi\|_\infty N^{-1/2}}{\sqrt{T-t_i}} + C\int_{t_i}^{t_{i+1}} \frac{dr}{(T-r)^{1-\gamma}\sqrt{r-t_i}} + \mathcal{E}(i) + C\sum_{j=i+1}^{N-1} \frac{\Theta_j\Delta_j}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j-t_i}}$$

Letting $U_i = \Theta_i$ and

$$W_i = C\|\Phi\|_\infty N^{-1/2}(T-t_i)^{-1/2} + C\int_{t_i}^{t_{i+1}} \frac{dr}{(T-r)^{1-\gamma}\sqrt{r-t_i}} + \mathcal{E}(i) =: \Gamma(i) + \Xi(i) + \mathcal{E}(i), \tag{1.7.35}$$

it follows from Lemma 1.8.6 that

$$\Theta_i \le CW_i + C\sum_{j=i+1}^{N-1} \frac{W_j\Delta_j}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j-t_i}} + C\sum_{j=i+1}^{N-1} \frac{\Theta_j\Delta_j}{(T-t_j)^{(1-\theta_L)/2}} \tag{1.7.36}$$

Therefore, using Lemma 1.8.7 in (1.7.33) and (1.7.34),

$$\|Z_{t_i} - Z_i\|_2 \leq W_i + C\sum_{j=i+1}^{N-1} \frac{W_j \Delta_j}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}}, \tag{1.7.37}$$

$$\|Y_{t_i} - Y_i\|_2 \leq CN^{-1/2} + C\sum_{j=i+1}^{N-1} \frac{\Psi(j)}{(T-t_j)^{(1-\theta_L)/2}} + C\sum_{j=i+1}^{N-1} \frac{W_j \Delta_j}{(T-t_j)^{(1-\theta_L)/2}}. \tag{1.7.38}$$

Let us consider the sum in the W terms. Firstly, remark that we only need consider the sums for $i < N-1$. Using the terminology of (1.7.35), Lemma 1.8.4 and Lemma 1.8.5,

$$\begin{aligned}\sum_{j=i+1}^{N-1} \frac{\Gamma(j)\Delta_j}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}} &= C\|\Phi\|_\infty N^{-1/2} \sum_{j=i+1}^{N-1} \frac{\Delta_j}{(T-t_j)^{1-\theta_L/2}\sqrt{t_j - t_i}} \\ &\leq C\|\Phi\|_\infty N^{-1/2}(T-t_i)^{(\theta_L - 1)/2},\end{aligned} \tag{1.7.39}$$

$$\begin{aligned}\sum_{j=i+1}^{N-1} \frac{\Xi(j)\Delta_j}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}} &\leq C\frac{\int_{t_{N-1}}^{T}(T-r)^{\gamma-1}(r-t_{N-1})^{-1/2}dr\Delta_{N-1}}{\Delta_{N-1}^{(1-\theta_L)/2}\sqrt{t_{N-1}-t_i}} \\ &\quad + C\sum_{j=i+1}^{N-2} \frac{\int_{t_j}^{t_{j+1}}(T-r)^{\gamma-1/2}(r-t_j)^{-1/2}dr\Delta_j}{(T-t_{j+1})^{1-\theta_L/2}\sqrt{t_j - t_i}} \\ &\leq \frac{C\Delta_{N-1}^{\gamma+\theta_L/2}}{\sqrt{t_{N-1}-t_i}} + C(T-t_i)^{(\theta_L-1)/2} \max_{0\leq i\leq N-1} \Delta_i^\gamma \\ &\leq CN^{-1}(t_{N-1}-t_i)^{-1/2} + CN^{-\gamma}(T-t_i)^{(\theta_L-1)/2}\end{aligned} \tag{1.7.40}$$

$$\begin{aligned}\sum_{j=i+1}^{N-1} \frac{\mathcal{E}(j)\Delta_j}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}} &= \sum_{j=i+1}^{N-1} \frac{\sum_{k=j+1}^{N-1} \frac{\Psi(k)}{(T-t_k)^{(1-\theta_L)/2}\sqrt{t_k - t_j}}\Delta_j}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}} \\ &= \sum_{k=i+2}^{N-1} \frac{\sum_{j=i+1}^{k-1} \frac{\Delta_j}{(T-t_j)^{1-\theta_L/2}\sqrt{t_j - t_i}}\Psi(k)}{(T-t_k)^{(1-\theta_L)/2}} \\ &\leq C\sum_{j=i+1}^{N-1} \frac{\Psi(j)}{(T-t_j)^{(1-\theta_L)/2}(t_j - t_i)^{(1-\theta_L)/2}} = C\mathcal{E}(i).\end{aligned} \tag{1.7.41}$$

where we have used

$$\sum_{j=i+1}^{N-2} \frac{\Delta_j}{(T-t_{j+1})^{1-\theta_L/2}\sqrt{t_j - t_i}} \leq C(T-t_i)^{(\theta_L-1)/2},$$

a result which can be proved analogously to Lemma 1.8.5. Using (1.7.39) - (1.7.41) implies that

$$\sum_{j=i+1}^{N-1} \frac{W_j \Delta_j}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}} \leq C(\|\Phi\|_\infty N^{-1/2} + N^{-\gamma})(T-t_i)^{(\theta_L-1)/2} + CN^{-1}(t_{N-1}-t_i)^{-1/2} + C\mathcal{E}(i). \tag{1.7.42}$$

Substituting this into (1.7.37) implies that

$$\sum_{i=0}^{N-1} \|Z_{t_i} - Z_i\|_2^2 \Delta_i \le \sum_{i=0}^{N-1} W_j^2 \Delta_i + C(\|\Phi\|_\infty^2 N^{-1} + N^{-2\gamma}) + CN^{-2}\ln(N) + C\sum_{i=0}^{N-1} \mathcal{E}(i)^2 \Delta_i$$
$$\le C(\|\Phi\|_\infty^2 N^{-1} + N^{-2\gamma}) + C\sum_{i=0}^{N-1} \Xi(i)^2 \Delta_i + C\sum_{i=0}^{N-1} \mathcal{E}(i)^2 \Delta_i \tag{1.7.43}$$

where we use the decomposition of W_j in (1.7.35). Using Lemma 1.8.4 and Lemma 1.8.5,

$$\begin{aligned}\sum_{i=0}^{N-1} \Xi(i)^2 \Delta_i &= \sum_{i=0}^{N-1} \Big(\int_{t_i}^{t_{i+1}} \frac{dr}{(T-r)^{1-\gamma}\sqrt{r-t_i}} \Big)^2 \Delta_i \\ &\le C\sum_{i=0}^{N-2} (T-t_{i+1})^{-1} \Big(\int_{t_i}^{t_{i+1}} \frac{dr}{(T-r)^{1/2-\gamma}\sqrt{r-t_i}} \Big)^2 \Delta_i + C\Delta_{N-1}^{2\gamma} \\ &\le CN^{-2\gamma}\ln(N) + CN^{-1}.\end{aligned}$$

Combining the above with the bound (1.7.19) of $\sum_i \mathcal{E}(i)^2\Delta_i$ in Proposition 1.7.5, it follows that

$$\sum_{i=0}^{N-1} \|Z_{t_i} - Z_i\|_2^2 \Delta_i \le CN^{-1} + CN^{-2\gamma}\ln(N) + CN^{(1-3\theta_L/4)/(2\gamma)-2}.$$

On the other hand, applying the Cauchy-Schwarz inequality and Hölder's inequality in (1.7.38), it follows that, for any i,

$$\|Y_{t_i} - Y_i\|_2^2 \le CN^{-1} + C\Big(\sum_{j=0}^{N-1} \frac{\Psi(j)}{(T-t_j)^{(1-\theta_L)/2}} \Big)^2 + C\sum_{j=0}^{N-1} W_j^2 \Delta_j.$$

We then use the bounds on $(\sum_{j=0}^{N-1} \Psi(j)(T-t_j)^{(\theta_L-1)/2})^2$ from Proposition 1.7.5 and the bounds on $\sum_{j=0}^{N-1} W_j^2\Delta_j$ computed above to complete the proof. □

Remark 1.7.7. The difference between the estimates of the approximation error of the ODP scheme 1.1.5, and $\mathcal{E}^M(N)$ for the Malliavin weights scheme (1.7.6) comes from the term $CN^{-2\gamma}\ln(N)$. In the special case $\theta_L = \alpha$ and $\theta_c = 1$ (whence $\gamma = \alpha$), which includes the case of the quadratic BSDE, it follows that the optimal convergence rate N^{-1} for the approximation error $\mathcal{E}^M(N)$ requires that α dominates $1/2 + \ln\ln(N)/\ln(N) > 1/2$. This is a more constrained set than that required by the ODP scheme, which may obtain the optimal convergence rate N^{-1} for α less than 1/2 - see Remark 1.6.7. However, it may be the case that the numerical resolution of the Malliavin weights scheme, where the conditional expectations are approximated by Monte Carlo least-squares regression, is more efficient. The study of the numerical algorithm based on the Malliavin weights scheme is the focus of Chapter 4.

1.8 Appendix

Lemma 1.8.1. *Let $f_s \in \mathbf{L}_2([0,T]\times\Omega)$. Then, for all $t \in [0,T]$, there exists a $\mathcal{B}([0,T])\otimes\mathcal{F}_t$-measurable processes F_t belonging to $L_2([0,T]\times\Omega)$ such that $(\omega,s)\mapsto F_t(s)$ is a version of $(\omega,s)\mapsto \mathbb{E}_t[f_s]$ and*

$$\mathbb{E}_t[\int_0^T f_s ds] = \int_0^T F_t(\cdot,s)ds \quad \textit{almost surely.}$$

Proof. Define the space

$$\mathcal{H} := \Big\{ f_s \in \mathbf{L}_2([0,T]\times\Omega) \ : \forall t\in[0,T]\ \exists \mathcal{B}([0,T])\otimes\mathcal{F}_t - \text{measurable process } F_t \in \mathbf{L}_2([0,T]\times\Omega)$$
$$\text{s.t } F_t(\cdot,s) = \mathbb{E}_t[f_s]\ dt\times\mathbb{P}-a.e. \text{ and } \mathbb{E}_t[\int_0^T f_s ds] = \int_0^T F_t(\cdot,s)ds \Big\}$$

The space $\mathcal{H}$ is a monotone linear vector space as in [RY99, Chapter 1, Theorem 2.2]; monotonicity follows from the monotone convergence theorem. The Lemma holds for the processes in the set

$$\mathcal{C} := \{f_s = \mathbf{1}_{A_0}\mathbf{1}_{\{0\}}(s) + \sum_i \mathbf{1}_{A_i}\mathbf{1}_{(t_i,t_{i+1}]}(s) : \quad (t_i) \text{ is any disjoint partition of } [0,T],\ A_i \in \mathcal{F}_T\}$$

with $F_t(s) = \mathbb{E}_t[\mathbf{1}_{A_0}]\mathbf{1}_{\{0\}}(s) + \sum_i \mathbb{E}_t[\mathbf{1}_{A_i}]\mathbf{1}_{(t_i,t_{i+1}]}(s)$. $\mathcal{C}$ is a subset of $\mathcal{H}$ that is closed under multiplication, whence $\mathcal{H}$ contains all bounded $\sigma(\mathcal{C})$-measurable processes by the Monotone Class Theorem. But $\sigma(\mathcal{C})$ is equal to $\mathcal{B}([0,T])\otimes\mathcal{F}_T$, so $\mathcal{H}$ contains all bounded $\mathcal{B}([0,T])\otimes\mathcal{F}_T$-measurable processes. The extension to processes in $\mathbf{L}_2([0,T]\times\Omega)$ is done by dominated convergence theorem. □

We need the following generalization of the a priori estimates [BDH+03, Proposition 3.2]:

Proposition 1.8.2. *Let k be an integer, and p be an integer greater than or equal to 2. Let $f : \Omega\times[0,T)\times(\mathbb{R}^k)^\top\times\mathbb{R}^{q\times k} \to (\mathbb{R}^k)^\top$ be $\mathcal{P}\times\mathcal{B}\big((\mathbb{R}^k)^\top\big)\otimes\mathcal{B}(\mathbb{R}^{q\times k})$-measurable, and ξ be an $(\mathbb{R}^k)^\top$-valued random variable in $\mathbf{L}_p(\mathcal{F}_T)$. Let $(f_t)_{t\in[0,T]}$ be non-negative, predictable process, $\mu \in \mathbf{L}_1([0,T];m)$ and $\lambda \in \mathbf{L}_2([0,T];m)$ be $\mathbb{R}$-valued non-negative. Additionally, assume that $\mathbb{E}[(\int_0^T f_t dt)^p] < \infty$. For any $(y_1,y_2)\in(\mathbb{R}^k)^2$, define the scalar product $(y_1,y_2) := \sum_{j=1}^k y_{1,j}y_{2,j}$ and assume that, for all $(t,y,z)\in[0,T)\times(\mathbb{R}^k)^\top\times\mathbb{R}^{k\times q}$, $(\omega,t,y,z)\mapsto f(\omega,t,y,z)$ satisfies*

$$\big(|y|^{-1}y\mathbf{1}_{|y|>0}, f(\omega,t,y,z)\big) \le f_t(\omega) + \mu_t|y| + \lambda_t|z| \quad \textit{almost surely.} \tag{1.8.1}$$

Let (Y,Z) be a solution to the $\big((\mathbb{R}^k)^\top,\mathbb{R}^{q\times k}\big)$-valued BSDE

$$Y_t = \xi + \int_t^T f(r,Y_r,Z_r)dr - \sum_{j=1}^q \int_t^T (Z_{j,r})^\top dW_{j,r}.$$

in the space $\mathcal{S}^p\times\mathcal{H}^p$, where $\mathcal{H}^p$ is the space of predictable processes X such that $\mathbb{E}[(\int_0^T |X_s|^2 ds)^{p/2}]$ is finite; Z_j denotes the j-th column of Z.

Then, there exists a constant C_p, depending only on p, such that, for any $\eta_t \ge \mu_t + \lambda_t^2/(p-1)$

in $L_1(\mathbb{R}; dt)$,

$$\mathbb{E}[\sup_t e^{p\int_0^t \eta_r dr}|Y_t|^p + (\int_0^T e^{2\int_0^t \eta_r dr}|Z_t|^2 dt)^{p/2}] \le C_p \mathbb{E}[e^{p\int_0^T \eta_r dr}|\xi|^p + (\int_0^T e^{\int_0^t \eta_r dr} f_t dt)^p].$$

Proof. Consider the processes $\tilde{Y}_t = e^{\int_0^t \eta_r dr} Y_t$ and $\tilde{Z}_t = e^{\int_0^t \eta_r dr} Z_t$. Then $(\tilde{Y}, \tilde{Z})$ satisfies a BSDE with terminal condition $\tilde{\xi} = e^{\int_0^T \eta_r dr}\xi$ and driver $\tilde{f}(t, y, z) = e^{\int_0^t \eta_r dr} f(t, e^{-\int_0^t \eta_r dr} y, e^{-\int_0^t \eta_r dr} z) - \eta_t y$. Moreover, for all $(t, y, z) \in [0, T) \times \mathbb{R}^k \times \mathbb{R}^{k\times q}$, $\tilde{f}(\omega, y, z)$ satisfies

$$\big(|y|^{-1} y \mathbf{1}_{|y|>0}, \tilde{f}(\omega, t, y, z)\big) \le \tilde{f}_t(\omega) + \tilde{\mu}_t |y| + \tilde{\lambda}_t |z| \quad \text{almost surely.}$$

with $\tilde{f}_t = e^{-\int_0^t \eta_r dr} f_t$, $\tilde{\mu}_t = \mu_t - \eta_t$, and $\tilde{\lambda}_t = \lambda_t$. The rest of the proof follows exactly as the proof of [BDH$^+$03, Proposition 3.2]. $\square$

From Proposition 1.8.2, we obtain the following generalization of [GM10, Lemma A.1]:

Lemma 1.8.3. *Let k be a positive integer, p be an integer greater than or equal to 2, and ξ be an $(\mathbb{R}^k)^\top$-valued random variable in $\mathbf{L}_2(\mathcal{F}_T)$. Let $(a_r)_r$, $(b_r)_r$ and $(c_{1,r}, \dots, c_{q,r})_r$ be progressively measurable processes, where a is $(\mathbb{R}^k)^\top$-valued, b is $\mathbb{R}^{k\times k}$-valued, and c_j is $\mathbb{R}^{k\times k}$-valued for each $j \in \{1, \dots, q\}$. If there exists $\theta \in (0, 1]$ and finite $L > 0$ such that for all $r \in [0, T)$, $|b_r| + \max_j |c_{j,r}| < L(T - r)^{(1-\theta)/2}$ almost surely, and if $\mathbb{E}(\int_0^T |a_r| dr)^2 < \infty$, then there exists a unique pair of processes (U, V) in $\mathcal{S}^2 \times \mathcal{H}^2$, where U takes values in $(\mathbb{R}^k)^\top$ and V takes values in $\mathbb{R}^{k\times q}$, solving the linear BSDE*

$$U_t = \xi + \int_t^T (a_r + b_r U_r + \sum_{j=1}^q (V_{j,r})^\top c_{j,r}) dr - \sum_{j=1}^q \int_t^T (V_{j,r})^\top dW_{j,r} \tag{1.8.2}$$

where V_j is the j-th column of V. Moreover, there is finite $C > 0$ depending only on θ, such that

$$\mathbb{E}[\sup_r |U_r|^2 + \int_0^T |V_r|^2 dr] \le C\|\xi\|_2 + (\int_0^T \|a_r\|_2 dr)^2 \tag{1.8.3}$$

Proof. Let (ϕ, ψ) be in $\mathcal{H}^2 \times \mathcal{H}^2$, and define the random function

$$f(r, y, z) = f(r) := a_r + b_r \phi_r + \sum_{j=1}^q \psi_{j,r} c_{j,r}.$$

The function f is progressively measurable and satisfies assumptions (H1)-(H5) of [BDH$^+$03, Section 4]. Since f takes no argument in (y, z), it is only necessary to check (H1): using Minkowski's inequality and the Cauchy-Schwarz inequality,

$$\begin{aligned}\mathbb{E}[(\int_0^T |f(r)| dr)^2]^{1/2} \le \mathbb{E}[(\int_0^T |a_r| dr)^2]^{1/2} + L(\int_0^T \mathbb{E}[|\phi_r|^2] dr)^{1/2} (\int_0^T \frac{dr}{(T-r)^{1-\theta}})^{1/2} \\ + L \sum_{j=1}^q (\int_0^T \mathbb{E}[|\psi_{j,r}|^2] dr)^{1/2} (\int_0^T \frac{dr}{(T-r)^{1-\theta}})^{1/2} < \infty.\end{aligned}$$

Thanks to [BDH+03, Theorem 4.2], there exists a unique solution (Y, Z) to the BSDE

$$Y_t = \xi + \int_t^T f(r)dr - \sum_{j=1}^{q} \int_t^T Z_{j,r} dW_r^j.$$

in $\mathcal{S}^2 \times \mathcal{H}^2$. The remainder of the proof of existence and uniqueness follows exactly as the proof of Corollary 1.4.3. To prove the bounds (1.8.3), observe that the driver $f(r)$ satisfies (1.8.1) from Proposition 1.8.2 with $f_r = a_r$ and $\lambda_r = \mu_r = C(T-r)^{(\theta-1)/2}$. □

Lemma 1.8.4. *The time grid* $\pi^{(\beta)} = \{0 = t_0 < \ldots < t_N = T \ : \ t_i = T - T(1-i/N)^{1/\beta}\}$ *with* $\beta \in (0,1]$ *satisfies*

$$\max_{0 \le i < N} \frac{\Delta_k}{(T-t_k)^{1-\theta}} \le \frac{T^\theta}{\beta} \frac{1}{N^{1 \wedge \frac{\theta}{\beta}}}, \tag{1.8.4}$$

$$\max_{0 \le i \le N-2} \frac{\Delta_k}{\Delta_{k+1}} \le \frac{1}{\beta}\Big(1 \vee (\frac{1}{2\beta})^{\frac{1}{\beta}-1}\Big), \tag{1.8.5}$$

for all $\theta \in (0,1]$.

The proof of this can be found in Lemma A.0.2.

Lemma 1.8.5. *Let* $\delta, \rho \in (0,1]$. *Then for* $B_{\delta,\rho} := \int_0^1 (1-r)^{\delta-1} r^{\rho-1} dr$, *for any* $0 \le s < t \le T$,

$$\int_t^s (s-r)^{\delta-1}(r-s)^{\rho-1} dr \le B_{\delta,\rho}(s-t)^{\delta+\rho-1}.$$

Moreover, on the time-grid $\pi^{(\beta)} = \{0 = t_0 < \ldots < t_N = T \ : \ t_i = T - T(1-i/N)^{1/\beta}\}$, *for any* $0 \le i < k \le N$,

$$\sum_{j=i+1}^{k-1} (t_k - t_j)^{\delta-1}(t_j - t_i)^{\rho-1} dr \le 2B_{\delta,\rho}(t_k - t_i)^{\delta+\rho-1}.$$

The proof of the integral part of Lemma 1.8.5 is done by changing variables. The proof of the sum part of Lemma 1.8.5 is exactly the same as the proof of Lemma 4.3.1 in Section 4.5.1 later. We note that, in the context of the notation of that Chapter, R_π is equal to 1, because the time-grid $\pi^{(\beta)}$ has the property $\Delta_{i+1} \le \Delta_i$ for all i.

Lemma 1.8.6. *Let* $\delta \ge 0, \rho > 0$ *and* $t \in [0,T)$. *Suppose that, for a positive constant* C_u, *the finite positive real functions* $u : [t,T] \mapsto [0,\infty)$ *and* $w : [t,T] \mapsto [0,\infty)$ *satisfy*

$$u_t \le w_t + C_u \int_t^T \frac{u_r dr}{(T-r)^{\frac{1}{2}-\delta}(r-t)^{\frac{1}{2}-\rho}}. \tag{1.8.6}$$

Then, for constants $\mathcal{C}_{(1.8.7a)}$ *and* $\mathcal{C}_{(1.8.7b)}$ *depending only on* C_u, T, δ *and* ρ,

$$u_t \le \mathcal{C}_{(1.8.7a)} w_t + \mathcal{C}_{(1.8.7a)} \int_t^T \frac{w_r dr}{(T-r)^{\frac{1}{2}-\delta}(r-t)^{\frac{1}{2}-\rho}} + \mathcal{C}_{(1.8.7b)} \int_t^T \frac{u_r dr}{(T-r)^{\frac{1}{2}-\delta}}. \tag{1.8.7}$$

Moreover, on the time-grid $\pi^{(\beta)} = \{0 = t_0 < \ldots < t_N = T \ : \ t_i = T - T(1-i/N)^{1/\beta}\}$, *suppose* *that the real functions* $U : \pi^{(\beta)} \mapsto [0,\infty)$ *and* $W : \pi^{(\beta)} \mapsto [0,\infty)$ *satify*

$$U_i \le W_i + C_u \sum_{j=i+1}^{N-1} \frac{U_j \Delta_j}{(T-t_j)^{\frac{1}{2}-\delta}(t_j-t_i)^{\frac{1}{2}-\rho}} \tag{1.8.8}$$

for all $i \in \{0,\ldots,N-1\}$. *It follows that*

$$U_i \le 2\mathcal{C}_{(1.8.7a)} W_i + 2\mathcal{C}_{(1.8.7a)} \sum_{j=i+1}^{N-1} \frac{W_j \Delta_j}{(T-t_j)^{\frac{1}{2}-\delta}(t_j-t_i)^{\frac{1}{2}-\rho}} + 2\mathcal{C}_{(1.8.7b)} \sum_{j=i+1}^{N-1} \frac{U_j \Delta_j}{(T-t_j)^{\frac{1}{2}-\delta}}$$

for all $i \in \{0,\ldots,N-1\}$.

The proof of Lemma 1.8.6 is analogous to the proof of Lemma 4.3.2 in Chapter 4. The integral part is proved in the same way as the sum part, with, of course, integrals replacing the sums.

Lemma 1.8.7. *Let* $\delta \ge 0, \rho > 0$ *and* $t \in [0,T)$. *Suppose that the finite positive real functions* $u : [t,T] \mapsto [0,\infty)$ *and* $w : [t,T] \mapsto [0,\infty)$ *satisfy* (1.8.7) *for some positive constants* $\mathcal{C}_{(1.8.7a)}$ *and* $\mathcal{C}_{(1.8.7b)}$. *Then, for* $\nu > 0$, *there is a positive constant* $\mathcal{C}^{(\nu)}$ *(depending only on* $\mathcal{C}_{(1.8.7a)}, \mathcal{C}_{(1.8.7b)}, T, \delta, \rho, \nu$*) such that*

$$\int_t^T \frac{u_r dr}{(T-r)^{\frac{1}{2}-\delta}(r-t)^{1-\nu}} \le \mathcal{C}^{(\nu)} \int_t^T \frac{w_r dr}{(T-r)^{\frac{1}{2}-\delta}(r-t)^{1-\nu}} \tag{1.8.9}$$

Moreover, on the time-grid $\pi^{(\beta)} = \{0 = t_0 < \ldots < t_N = T \ : \ t_i = T - T(1-i/N)^{1/\beta}\}$, *suppose that the real functions* $U : \pi^{(\beta)} \mapsto [0,\infty)$ *and* $W : \pi^{(\beta)} \mapsto [0,\infty)$ *satify* (1.8.8) *for all* $i \in \{0,\ldots,N-1\}$. *It follows that*

$$\sum_{j=i+}^{N-1} \frac{U_j \Delta_j}{(T-t_j)^{\frac{1}{2}-\delta}(t_j-t_i)^{1-\nu}} \le 2\mathcal{C}^{(\nu)} \sum_{j=i+1}^{N-1} \frac{W_j \Delta_j}{(T-t_j)^{\frac{1}{2}-\delta}(t_j-t_i)^{1-\nu}}$$

for all $i \in \{0,\ldots,N-1\}$.

The proof of Lemma 1.8.7 is analogous to the proof of Lemma 4.3.3 in Chapter 4. The integral part is proved in the same way as the sum part, with, of course, integrals replacing the sums.

2 Approximation of discrete BSDE using least-squares regression

2.1 Introduction

Framework. Let $T > 0$ be a fixed terminal time and W be a q-dimensional ($q \geq 1$) Brownian motion defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where the filtration $(\mathcal{F}_t)_{0 \leq t \leq T}$ satisfies the usual hypotheses; the filtration may be larger than that generated by W. We are given a deterministic time grid $\pi := \{0 = t_0 < \ldots < t_N = T\}$ for the interval $[0, T]$, whose $(i+1)$-th time-step is denoted $\Delta_i = t_{i+1} - t_i$ and mesh size is defined by $|\pi| := \max_{0 \leq i < N} \Delta_i \leq T$. The $(i+1)$-th Brownian motion increment is defined by $\Delta W_i := W_{t_{i+1}} - W_{t_i}$.

In this chapter, we deal with the numerical resolution of (Y, Z), a *discrete* BSDE with data $(\xi, f_i(y, z))$, which is generated by

$$\begin{cases} Y_N &= \xi, \quad Y_i = \mathbb{E}_i\left(Y_{i+1} + f_i(Y_{i+1}, Z_i)\Delta_i\right), \ 0 \leq i < N, \\ \Delta_i Z_i &= \mathbb{E}_i\left(Y_{i+1}\Delta W_i^{\top}\right), \quad 0 \leq i < N \end{cases} \tag{2.1.1}$$

where $\mathbb{E}_i(\cdot) := \mathbb{E}(\cdot|\mathcal{F}_{t_i})$, $\top$ denotes the transpose operator and

- ξ is a given $\mathcal{F}_T$-measurable random variable in $\mathbf{L}_2$,
- $Y := (Y_i)_{0 \leq i \leq N}$ is a scalar process, $Z := (Z_i)_{0 \leq i < N}$ is $\mathbb{R}^q$-valued process (as a row vector),
- for each i, the so-called driver $(\omega, y, z) \mapsto f_i(y, z)$ is $\mathcal{F}_{t_i} \otimes \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}^q)$-measurable.

Equation (2.1.1) is a backward Dynamic Programming (DP for short) equation, which is solved at $i = N-1$ by first evaluating Z_{N-1} using $Y_N = \xi$, then Y_{N-1} using Y_N and Z_{N-1}, and then iterating these evaluations until $i = 0$.

Application. Equation (2.1.1) appears naturally when approximating a continuous-time BSDE by a discrete-time process along the time grid π. The continuous time BSDE may be a generalized BSDE of the form

$$Y_t = \xi + \int_t^T f(s, Y_s, Z_s)ds - \int_t^T Z_s dW_s - (L_T - L_t), \tag{2.1.2}$$

where L is a martingale orthogonal to W. The presence of L occurs, for example, when $\xi = \Phi(X_T)$ where X is a jump-diffusion process driven by W and a Poisson random measure. In this context, and for $f(s, \omega, y, z) = f(s, X_s(\omega), y, z)$, it is shown in [LGW06, Theorem 1] that the discrete time process $(Y_i, Z_i)_{0 \leq i < N}$ generated by (2.1.1) converges to (Y, Z), in suitable $\mathbf{L}_2$-spaces, as the mesh size $|\pi|$ goes to 0. Note that the general formulation (2.1.1) using conditional expectations w.r.t. $\mathcal{F}_{t_i}$ allows for path-dependent drivers/terminal conditions: one can take $\xi = \Phi(X_{t_1}, \ldots, X_{t_N})$ (see [GGG12]) or $\xi = \Phi(X_T, \int_0^T X_t dt)$ for a diffusion process X (see [GLW05]).

DP equation (2.1.1) is written in an explicit form because the driver f_i depends on Y_{i+1}. Discrete BSDEs have traditionally been studied in implicit form, i.e. where f_i depends on Y_i. The explicit and implicit schemes usually give, to the best of our knowledge, the same rate of convergence as $N \to +\infty$ for the discretization error (the error incurred by approximating the continuous BSDE by a discrete process) in an appropriate $\mathbf{L}_2$-space: compare the results of [BT04] (implicit scheme) to [LGW06] (explicit scheme).

A lot of attention - [Zha04][BT04][GM10] and Chapter 1 of this thesis among others - has been paid to the analysis of the discretization error. This is not the focus of the current chapter; rather, we focus on the numerical resolution of the DP equations (2.1.1), allowing $N \to +\infty$.

Numerical approximation. One has to approximate the conditional expectations in (2.1.1) in order to have a fully implementable scheme. This is because one cannot, in general, calculate the conditional expectation explicitly. Over the last ten years, various different approaches have been developed to do this - see the introduction of [GL10] for an overview - but very few papers [BT04][LGW06][Mos10] have tackled the global error analysis. In the current chapter, we follow the *empirical regression approach* presented in [LGW06] and estimate the global error that this method incurs in the approximation of (Y, Z). Suppose that

$$(Y_i, Z_i) := (y_i(X_i), z_i(X_i)) \tag{2.1.3}$$

for some (unknown but deterministic) measurable functions $y_i(\cdot), z_i(\cdot)$ and a d-dimensional explanatory process $X := (X_i)_{0 \le i \le N}$ (in the jump-diffusion example above, X would be the Euler approximation at times $(t_i)_i$). Since each conditional expectation $\mathbb{E}_i(\cdot)$ can be viewed as solution of a least-squares problem in $\mathbf{L}_2(\mathbb{P})$, the functions $y_i(\cdot)$ and $z_i(\cdot)$ are then approximated using a finite-dimensional approximation. The coefficients of this approximation are computed using empirical least-squares regression [GKKW02, Chapters 10-11-12] using M simulations of the paths of the explanatory variable X. The use of such an empirical regression scheme is supported by two important features: first, it requires as an input only independent paths of the explanatory process X and of the Brownian motion W; second, using distribution-free tools [GKKW02], one may achieve model-independent error estimates related to the statistical error. These robust estimates (Theorem 2.4.5) are presumably too conservative; on the other hand, the estimates allow the error analysis to be applied to very general probability spaces, because we make very few assumptions on the explanatory process X: see Section 2.2.

The resulting global error is known to be very difficult to analyze, because all regression problems are stochastically dependent through the DP equation; moreover, the numerical parameters (the time-grid π, functions basis used for the finite-dimensional approximations of $y_i(\cdot), z_i(\cdot)$ and the number of simulations M) play multiple, often contradictory, roles in the convergence, and it is crucial to find the right trade-off between them. We achieve the global error analysis in Theorem 2.4.4 which is our main results. We then apply this result to optimize the numerical parameters needed for a given accuracy in the asymptotics $N \to +\infty$, see Subsection 2.4.3. For the reader interested in empirical analysis, we refer to [Mos10][Ric10].

2.1.1 Our contributions

In this chapter, we introduce an algorithm using the Multi step-forward Dynamic Programming (MDP for short) equation given by

$$\begin{cases} Y_i & = \mathbb{E}_i \left(\xi + \sum_{k=i}^{N-1} f_k(Y_{k+1}, Z_k)\Delta_k \right), \\ \Delta_i Z_i & = \mathbb{E}_i \left([\xi + \sum_{k=i+1}^{N-1} f_k(Y_{k+1}, Z_k)\Delta_k]\Delta W_i^\top \right). \end{cases} \tag{2.1.4}$$

Equation (2.1.4) is inspired by the algorithm of [BD07], but we notice that, unlike that work, Picard iterations are not used in our scheme. Because of the tower property of conditional expectations, definitions (2.1.1) and (2.1.4) coincide. When Least-Squares approximations of the conditional expectations are incorporated in the MDP equation, this gives a so-called *LSMDP scheme* and we provide an equivalent result to [BD07, Theorem 11] in Theorem 2.3.6, which shows that our LSMDP scheme gives raise to a smaller propagation of errors than the the One step-forward Dynamic Programming (ODP for short) equation (2.1.1); in this sense, we suggest that the Picard iterations of [BD07, Mos10] are unnecessary. This is good news because it greatly simplifies the algorithm and its analysis without deteriorating the estimates.
Moreover, we provide several important relaxations to the traditional assumptions of numerical schemes for BSDEs [LGW06][BD07][Mos10]:

(a) We allow the driver to satisfy a weaker, local Lipschitz condition; see ($\mathbf{A_F}$-**i**)below. This allows the results of this chapter to be applied to a wider range of approximation problems for BSDEs, including the important class of quadratic BSDEs with bounded, Hölder continuous terminal conditions. To the best of our knowledge, a detailed error analysis for numerical schemes for this class of BSDEs has not before been performed.

(b) We allow the driver to satisfy a weaker, local bound at $(y, z) = (0, 0)$; see ($\mathbf{A_F}$-**ii**)below. This allows the results of this chapter to be applied to a particular proxy technique. This is similar to the martingale basis method of Bender-Steiner [BS12]. The authors give an example of where this method performs much better than the traditional One-step scheme, but are unable to analyze the error in a general setting; the results of this chapter now allow this analysis.

We defer to Section 2.2 for a detailed discussion of the assumptions and the new applications of our results refered to above.

We derive a full error analysis for the LSMDP scheme, including the effect of statistical errors (finite number of Monte Carlo simulations); see Theorem 2.4.4. These estimates are obtained by exploiting stability inequalities - Section 2.3.1 - for discrete BSDEs. We demonstrate how higher orders of smoothness of the Markov functions y_i and z_i defined in (2.1.3) lead to improvements in the error-computational work trade-off. We also demonstrate that MDP has a better error-computational time trade off than ODP in the context of this theoretical analysis; see Subsection 2.4.3. Moreover, we do not require Lipschitz continuity assumptions on the Markov functions y_i to obtain our estimates.

What is particularly interesting is that the computational efficiency computed in Section 2.4.3 is of the same order whether we take local or global Lipschitz continuity of the driver. The local continuity corresponds to the quadratic BSDE problem. In this sense, we obtain the optimal rates of convergence for the quadratic BSDE. The truncation method of [IDR10, Section 6], for example, requires the use of a truncation of the driver by a smooth projection of the z component onto the open ball or radius $\bar{R}$. This will reduce the quadratic driver to a Lipschitz continuous driver whose Lipschitz coefficent depends on $\bar{R}$, and they require that $\bar{R}$ be large so that their approximation error is low [IDR10, Theorem 6.2]. As we will see in Theorems 2.4.4 and 2.4.5, the efficiency of the LSMDP algorithm depends on the absolute bounds of y_i and z_i, which depend exponentially on the Lipschitz constant, see Proposition 2.3.8. This presents a substantial reduction of the numerical efficiency of the numerical scheme. This phenomenon has already been observed in the introduction of [Ric11], where a nice heuristic argument is provided. Another alternative approach is to apply he Cole-Hopf transform [IDRZ10], which transforms a quadratic BSDE into a Lipschitz BSDE. Numerical resolution of the transformed problem leads to good convergence rates. However, the Cole-Hopf transform can only be applied to certain quadratic drivers. Our method accommodates general quadratic drivers. The method of [Ric11] the use of an irregular time-grid to attain an optimal convergence rate. Although our algorithm accommodates this special time grid, it is also able to accommodate the time grids of Chapter 1, where the time points take the form $t_i = T - T(1 - i/N)^{1/\theta_\pi}$. It was shown in Chapter 1 that these time grids improve the rate of convergence of the time-discrete scheme to the continuous BSDE.

We allow the time grid π to be non uniform; see ($\mathbf{A_F}$**-iii**). Indeed, to reduce the discretization error for BSDE with irregular terminal conditions $\xi = \Phi(X_T)$, it has been recently proposed in [GM10] to choose nonuniform grids: the grid points are more concentrated close to the terminal time T in order to compensate the lack of regularity of Φ. The results of this chapter can be applied to the time-grids of [GM10]. Similar results are obtained for for path dependent ξ in [GGG12].

We remark that our analysis in Section 2.4 bears similarity to that of [Mos10, Chapter 3] at first sight. We want to briefly compare our results with that work. The assumptions of [Mos10]are stronger: there is a uniform Lipschitz condition and uniform bound on the driver. This does not treat the extensions of the scheme discussed above. Weakening the assumptions leads to problems in the analysis that cannot be treated in a trivial way, and we provide a careful treatment of these issues. See the proofs of theorems 2.4.4 and 2.4.5 for details. Indeed, that the estimates are in the end similar is, to us, rather astonishing given the broadening of the spectrum of problems that one can treat and previous difficulties that one had with these problems (see above). In addition, our method does not require the use of Picard iterations. The estimates of Theorem 2.4.4 and 2.4.5 are in terms of the discrete or continuous time BSDE. This allows the use of the smoothness of the true BSDE to get finer estimates. In contrast, the estimates of [Mos10, Theorem 3.4.1] are expressed in terms of the Picard iterations, and it is not clear that one can take advantage of the additional smoothness in the same way.

Organization of the chapter. In the remainder of this section, we define notation used throughout the chapter. In Section 2.2, we state our working assumptions and give several examples to show how these assumptions are useful for approximating a wide variety of continuous-time BSDEs. In Section 2.3, we establish stability estimates for discrete BSDEs, and apply them to derive tight pointwise and $\mathbf{L}_2$-estimates for (Y, Z). We define the MDP-based scheme and we analyze the $\mathbf{L}_2$-error incurred when conditional expectations are approximated by projections on closed convex subsets of $\mathbf{L}_2$. This allows comparison between ODP and MDP-based schemes. Finally, we determine boundedness and smoothness properties of discrete BSDEs under some additional assumptions. In Section 2.4, the projections are computed using M independent simulations of the explanatory process X: it defines the LSMDP scheme. The global error is stated in Theorems 2.4.4 and 2.4.5. The rest of the section is devoted to (long and technical) proofs. A discussion related to algorithm complexity is given in Subsection 2.4.3. In particular, we show how higher order of smoothness of the Markov functions y_i and z_i leads to an improved error-computational time trade-off, and compare this to the results for the ODP. Some intermediate results are detailed in the Appendix.

Further notation.

- $|x|$ stands for the Euclidean norm of the vector x.
- $|U|_{\mathbf{L}_p} = (\mathbb{E}|U|^p)^{\frac{1}{p}}$ stands for the $\mathbf{L}_p(\mathbb{P})$-norm ($p \geq 1$) of a random variable U. To indicate that U is additionally measurable w.r.t. the σ-algebra $\mathcal{Q}$, we may write $U \in \mathbf{L}_p(\mathcal{Q}, \mathbb{P})$.
- We reserve the letter $\gamma := (\gamma_0, \ldots, \gamma_{N-1}) \in \mathbb{R}_+^N$ for the parameter appearing in the weighted $\mathbf{L}_2$-norms below. Moreover for a given γ, we set $\Gamma_i := \prod_{k=0}^{i-1}(1 + \gamma_k \Delta_k)$ for $0 \leq i < N$ (with the usual convention $\prod_{k=0}^{-1} \cdots = 1$).
- For a $q(> 1)$-dimensional process $U = (U_i)_{0 \leq i \leq N}$, its l-th component is denoted by $U_l = (U_{l,i})_{0 \leq i \leq N}$.

2.2 Standing assumptions and applicability to practical continuous-time problems

In this section, we give the standing assumptions for this chapter. These assumptions are more general than in previous numerical schemes for BSDEs, and we outline several examples to demonstrate how these more general assumptions lead extended applicability of the results of this chapter to practical continuous-time BSDE problems.

The standing assumptions are separated into two parts: the first set consists of the minimal assumptions necessary for basic results of Section 2.3, and the second set consists of the *Markovian* assumptions necessary for Section 2.4. The minimal assumptions used in this chapter are that the terminal condition ξ is square integrable and the driver is *locally* Lipschitz continuous in the sense that the Lipschitz constant depends on t_i. To be more precise:

($\mathbf{A}_\xi$) ξ is in $\mathbf{L}_2(\mathcal{F}_T, \mathbb{P})$.

($\mathbf{A_F}$) i) $(\omega, y, z) \mapsto f_i(y, z)$ is $\mathcal{F}_{t_i} \otimes \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}^q)$-measurable for every $i < N$, and there exist deterministic parameters $\theta_L \in (0, 1]$ and $L_f \in [0, +\infty)$ such that

$$|f_i(y, z) - f_i(y', z')| \leq \frac{L_f}{(T - t_i)^{(1-\theta_L)/2}}(|y - y'| + |z - z'|), \tag{2.2.1}$$

for any $(y, y', z, z') \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^q \times \mathbb{R}^q$.

ii) There exist deterministic parameters $\theta_c \in (0,1]$ and $C_f \in [0,+\infty)$ such that

$$|f_i(0,0)| \le \frac{C_f}{(T-t_i)^{1-\theta_c}}, \quad \forall 0 \le i < N. \tag{2.2.2}$$

iii) The time-grids $\pi := \{0 = t_0 < \ldots < t_N = T\}$ are such that

$$C_\pi = \sup_{k<N} \frac{\Delta_k}{(T-t_k)^{1-\theta_L}} \to 0 \quad \text{as } N \to +\infty, \tag{2.2.3}$$

$$\limsup_{N\to\infty} R_\pi < +\infty, \quad \text{where } R_\pi = \sup_{0\le k\le N-2} \frac{\Delta_k}{\Delta_{k+1}}. \tag{2.2.4}$$

Under $(\mathbf{A}_\xi)$and $(\mathbf{A_F}\text{-}\mathbf{i}\text{-}\mathbf{ii})$, it is straightforward to check from (2.1.1) that $(Y_i)_{0\le i\le N}$ and $(Z_i)_{0\le i<N}$ are well defined and belong to $\mathbf{L}_2$ (see Proposition 2.3.3 for tight estimates).

When analyzing the influence of M in Section 2.4, we reinforce the basic assumptions with the following set of *Markovian* assumptions:

$(\mathbf{A_X})$ X is a Markov chain in $\mathbb{R}^d$ $(1 \le d < +\infty)$ adapted to $(\mathcal{F}_{t_i})_i$.

$(\mathbf{A}'_\xi)$ i) ξ is a bounded $\mathcal{F}_T$-measurable random variable; we set $C_\xi := \mathbb{P} - \operatorname{ess\,sup}_\omega |\xi(\omega)| < +\infty$.

ii) ξ is of form $\xi = \Phi(X_N)$ for a measurable function Φ.

$(\mathbf{A}'_\mathbf{F})$ For every $i < N$, the driver is of the form $f_i(y,z) = f_i(X_i, y, z)$ where $(x,y,z) \mapsto f_i(x,y,z)$ is $\mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}^q)$-measurable and $(\mathbf{A_F})$is satisfied.

These give us a Markov representation for solutions of the discrete BSDEs: for all $k < N$, there exist measurable, deterministic functions $y_k : \mathbb{R}^d \to \mathbb{R}$ and $z_k : \mathbb{R}^d \to \mathbb{R}^q$ such that $Y_k = y_k(X_k)$ and $Z_k = z_k(X_k)$ holds almost surely. Indeed, taking the Markov chain $(X_i^{k,x})_{i\ge k}$ started at t_k with value $x \in \mathbb{R}^d$ with the same transition probabilities as $(X_k)_k$ yields, by induction, $\Delta_k z_k(x) = \mathbb{E}[\Delta W_k^\top y_{k+1}(X_{k+1}^{k,x})]$ and $y_k(x) = \mathbb{E}[y_{k+1}(X_{k+1}^{k,x}) + f_k(x, y_{k+1}(X_{k+1}^{k,x}), z_k(x))\Delta_k]$; see the proof of Lemma 2.4.3 for details.

We emphasize that we do not make any further assumptions on X - no non-degeneracy condition, no specific distributions, etc; our error estimates are *model-free* in this sense. This lends flexibility and robustness to the empirical least-squares regression scheme.

The assumptions above derive from particular continuous time settings; see below. This means that it is natural for us to assume that the constants θ_L, L_f, θ_c, and C_f are time-grid independent. This assumption simplifies the complexity analysis in Subsection 2.4.3.

At first glance, the boundedness assumption $(\mathbf{A}'_\xi\text{-}\mathbf{i})$appears to be a serious restriction of our scheme. Indeed, $(\mathbf{A}_\xi)$is the minimal assumption to ensure the existence of a continuous-time BSDE in $\mathbf{L}_2$-spaces [EKPQ97]. The *raison d'être* of $(\mathbf{A}'_\xi\text{-}\mathbf{i})$is to derive robust estimates for the global error (see Theorem 2.4.4) using the tools of nonparametric regression [GKKW02]. On the other hand, $\xi_n = -n \vee \xi \wedge n$ $(n \ge 0)$ defines a sequence of bounded approximations of ξ and by $\mathbf{L}_2$-stability results on continuous-time BSDEs (see [EKPQ97, Proposition 2.1] for instance), the truncation error converges to 0 as $n \to +\infty$. Since in our global error estimates we keep track on the dependence on C_ξ, it would be a priori possible to let this upper bound go appropriately quickly to infinity, while maintaining a converging scheme.

Assumptions ($\mathbf{A_F}$-**i-ii**)may be surprising because they extend the usual global Lipschitz continuity conditions in an unusual way. Globally Lipschitz drivers are related to the case $\theta_L = 1$, and $\theta_c = 1$ describes the usual situation where drivers are uniformly (in time) bounded at $(y, z) = (0, 0)$. The singularity at the terminal time allows us to extend the applicability of our numerical scheme to include a wider class of continuous-time Markovian BSDE related to a $\mathbb{R}^d$-valued Brownian diffusion process $(X_t)_{0\le t\le T}$. We outline two canonical examples that motivate ($\mathbf{A_F}$-**i-ii**). Take $\xi = \Phi(X_T)$ and $f(t, \omega, y, z) = f(t, X_t(\omega), y, z)$. For simplicity, assume $q = d$ and that the coefficients of X are smooth and bounded and that its diffusion coefficient $\sigma(t, x)$ satisfies a uniform ellipticity condition. We denote by $\mathcal{L}$ the infinitesimal generator of X.

Quadratic BSDEs. Consider a quadratic growth driver satisfying

$$|f(t, x, y, z)| \le c\ (1 + |y| + |z|^2),$$
$$|f(t, x, y, z) - f(t, x, y', z')| \le c\ (1 + |z| + |z'|)(|x - x'| + |y - y'| + |z - z'|)$$

for any $(t, x, x', y, y', z, z') \in [0, T]\times\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}\times\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}^d$ and for a given constant $c \ge 0$. Assume additionally that the terminal function Φ is Hölder continuous and bounded. Then [DG06, Theorem 2.1] yields that the continuous-time BSDE is given by $Y_t = u(t, X_t)$ and $Z_t = \nabla u(t, X_t)\sigma(t, X_t)$ where u solves the semi-linear PDE $\partial_t u(t, x) + \mathcal{L}u(t, x) + f(t, x, u(t, x), \nabla u(t, x)\sigma(t, x)) = 0$ with $u(T, x) = \Phi(x)$. Moreover, there exist constants $\theta \in (0, 1]$ and $C_u \in \mathbb{R}^+$ such that

$$(T - t)^{(1-\theta)/2}|\nabla u(t, x)\sigma(t, x)| \le C_u, \quad \forall (t, x) \in [0, T) \times \mathbb{R}^d.$$

Now, set $\varphi_t : \zeta \in \mathbb{R} \mapsto \varphi_t(\zeta) = \operatorname{sign}(\zeta) \min\big(|\zeta|, \frac{C_u}{(T-t)^{(1-\theta)/2}}\big)$ and define the new driver $\bar{f}(t, x, y, z) := f\big(t, x, y, \varphi_t(z_1) \ldots, \varphi_t(z_d)\big)$. Observe that $\bar{f}(t, X_t, Y_t, Z_t) = f(t, X_t, Y_t, Z_t)$, thus it is equivalent to solve the BSDE with driver f or $\bar{f}$. Notice also that $\varphi_t(\cdot)$ is 1-Lipschitz continuous and bounded by $\frac{C_u}{(T-t)^{(1-\theta)/2}}$, hence $f_i(y, z) := \bar{f}(t_i, X_{t_i}, y, z)$ satisfies ($\mathbf{A_F}$-**i-ii**)with $C_f = c$, $\theta_c = 1$, $L_f = c(T^{(1-\theta)/2} + 2\sqrt{d}C_u)$, $\theta_L = \theta$. In Chapter 1, it was proved that the exponent θ is equal to the Hölder exponent of Φ, and it is possible to obtain an explicit estimate of the constant C_u.

Using proxys for numerical stability. Consider a standard Lipschitz driver f. Assume that we know by *expertise* that the solution $(Y_t, Z_t)_t$ is expected to be close to $(v(t, X_t), \nabla v(t, X_t)\sigma(t, X_t))_t$, where v is the explicit solution to a linear parabolic equation $\partial_t v(t, x) + \tilde{\mathcal{L}}v(t, x) + \tilde{f}(t, x) = 0$; the diffusion process associated to $\tilde{\mathcal{L}}$, the terminal condition and the driver may have changed to produce an analytical solution. v is called *proxy* in [BGM09]. It is then natural to numerically compute the residual $(Y^0_t, Z^0_t) := (Y_t - v(t, X_t), Z_t - \nabla v(t, X_t)\sigma(t, X_t))$. It solves a BSDE with terminal function $\Phi(.) - v(T, .)$ and driver

$$f^0(t, x, y, z) := f(t, x, y + v(t, x), z + \nabla v(t, x)\sigma(t, x)) - \tilde{f}(t, x) + (\mathcal{L} - \tilde{\mathcal{L}})v(t, x).$$

The new driver f^0 is uniformly Lipschitz w.r.t. y and z, so ($\mathbf{A_F}$-**i**)is satisfied with $\theta_L = 1$. If $v(T, .)$ is θ-Hölder continuous ($\theta \in (0, 1]$), then usual PDE estimates on the parabolic operator $\tilde{\mathcal{L}}$ give $(T - t)^{(\frac{k-\theta}{2})_+}|D^k_x v(t, x)| \le C_v$ $(k = 0, 1, 2)$, from which ($\mathbf{A_F}$-**ii**)is derived with $\theta_c = \theta/2$. To conclude this example, we mention that in the case $\tilde{\mathcal{L}} = \mathcal{L}$, $v(T, .) = \Phi(.)$ and $\tilde{f} = 0$, it is proved in [GM10] that the $\mathbf{L}_2$-time-regularity of (Y^0, Z^0) is usually more well-behaved than that of (Y, Z), suggesting that the discretization error from the DP equation for (Y^0, Z^0) would be smaller.

Assumption ($\mathbf{A_F}$-**iii**)is used to derive stability results for discrete BSDEs (see Proposition 2.3.3) and for the numerical schemes (see Theorems 2.3.6 and 2.4.4) as the number N of grid times becomes large. If $\theta_L = 1$, the condition (2.2.3) is equivalent to $|\pi| \to 0$. If $\theta_L \in (0,1)$ and π is a time-grid with higher concentration at T as in [GM10] (i.e. $t_k = T - T(1-k/N)^{1/\theta_\pi}$ with $\theta_\pi \in (0,1]$), then one easily checks conditions (2.2.3) and (2.2.4) hold (see Lemma A.0.2):

$$C_\pi \le \frac{T^{\theta_L}}{\theta_\pi}\frac{1}{N^{1\wedge\frac{\theta_L}{\theta_\pi}}}, \qquad R_\pi \le \frac{1}{\theta_\pi}\left(1 \vee \big(\frac{1}{2\theta_\pi}\big)^{\frac{1}{\theta_\pi}-1}\right).$$

Hence, (2.2.3) and (2.2.4) hold true whatever the value θ_π is. This shows that in most usual situations $|\pi| \to 0$ implies ($\mathbf{A_F}$-**iii**).

2.3 ODP scheme vs. MDP scheme

The aim of this section is threefold. Firstly, we use the minimal assumptions ($\mathbf{A}_\xi$)and ($\mathbf{A_F}$)to determine a priori stability estimates for discrete BSDEs. Secondly, we use the stability results to show that the MDP scheme combined with $\mathbf{L}_2$-projection yields a smaller error than the ODP scheme. In doing so, we revisit the results of [BD07], but avoid the Picard iterations of their scheme. This also serves as a warm-up to Section 2.4, where the a priori stability results also play a crucial role. Thirdly, we demonstrate how slightly stronger assumptions yield time uniform almost sure bounds on the solutions of the BSDEs, and smoothness properties in the case of Markovian BSDEs; these properties are extremely useful in Section 2.4.

2.3.1 General a priori estimates

Definition 2.3.1. *The truncation of the i-th Brownian increment at threshold* $R = [0, +\infty]$ *is defined by*

$$[\Delta W_i]_w = (-R\sqrt{\Delta_i} \vee \Delta W_{1,i} \wedge R\sqrt{\Delta_i}, \cdots, -R\sqrt{\Delta_i} \vee \Delta W_{q,i} \wedge R\sqrt{\Delta_i})^\top.$$

For $R = +\infty$, $[\Delta W_i]_w = \Delta W_i$. Replacing ΔW_i by $[\Delta W_i]_w$ has small impact in the DP equations (2.1.1), provided that R is large enough (see Proposition 2.3.9). On the other hand, taking finite R ensures that some quantities are a.s. bounded, which is crucial in our error analysis in Section 2.4.

In this subsection, we study the difference between two *discrete BSDEs* with truncated Brownian increments, $(Y_i^R, Z_i^R)_i$ and $(\bar Y_i^R, \bar Z_i^R)_i$, given by

$$\begin{cases} Y_i^R &= \mathbb{E}_i\left(\xi + \sum_{k=i}^{N-1} f_k(Y_{k+1}^R, Z_k^R)\Delta_k\right), \\ \Delta_i Z_i^R &= \mathbb{E}_i\left([\xi + \sum_{k=i+1}^{N-1} f_k(Y_{k+1}^R, Z_k^R)\Delta_k][\Delta W_i^\top]_w\right), \end{cases} \tag{2.3.1}$$

and similarly for $(\bar Y_i^R, \bar Z_i^R)_i$ with data $(\bar\xi, \bar f_i(y,z))$. The superscript R refers to the fact the Brownian increments are truncated at the threshold $R \in [0, +\infty]$.

We assume that ξ and $\bar\xi$ are both in $\mathbf{L}_2$ (assumption ($\mathbf{A}_\xi$)). We allow rather greater generality than afforded by ($\mathbf{A_F}$): firstly, the drivers $f_i(y,z)$ and $\bar f_i(y,z)$ are Lipschitz continuous w.r.t. (y,z)

and the dependence of their Lipschitz constant w.r.t. i is general; finally, we do not insist that the drivers be adapted. We will require the extension to non-adapted drivers later in Section 2.4, where we will apply these results to BSDEs with data dependent drivers. However, we assume that each $f_i(Y^R_{i+1}, Z^R_i)$ and $\bar{f}_i(\bar{Y}^R_{i+1}, \bar{Z}^R_i)$ are in $\mathbf{L}_2$, so that Y^R_i, Z^R_i, $\bar{Y}^R_i$, $\bar{Z}^R_i$ are also in $\mathbf{L}_2$ for any i. Using the tower property of conditional expectations, observe that

$$\begin{cases} Y^R_N &= \xi, \quad Y^R_i = \mathbb{E}_i\left(Y^R_{i+1} + f_i(Y^R_{i+1}, Z^R_i)\Delta_i\right), \\ \Delta_i Z^R_i &= \mathbb{E}_i\left(Y^R_{i+1}[\Delta W^\top_i]_w\right), \end{cases} \tag{2.3.2}$$

and similarly for $(\bar{Y}^R, \bar{Z}^R)$. We study the differences:

$$\Delta Y^R_i = Y^R_i - \bar{Y}^R_i, \qquad \Delta Z^R_i = Z^R_i - \bar{Z}^R_i$$

and we set

$$\Delta f_i = f_i(Y^R_{i+1}, Z^R_i) - \bar{f}_i(Y^R_{i+1}, Z^R_i), \qquad \Delta\xi = \xi - \bar{\xi}.$$

We shall use the following Lemma repeatedly:

Lemma 2.3.2 (Local estimates)**.** *For $0 \le i \le N-1$, assume that $\bar{f}_i$ is Lipschitz w.r.t. y and z, with a Lipschitz constant equal to $L_{\bar{f}_i} \in \mathbb{R}^+$. For any $R \in [0, +\infty]$, $\Delta_i \le T$ and $\gamma_i > 0$ satisfying $6q(\Delta_i + \frac{1}{\gamma_i})L^2_{\bar{f}_i} \le 1$, we have*

$$|\Delta Y^R_i|^2 \le (1 + (\gamma_i + \frac{1}{2})\Delta_i)\mathbb{E}_i(|\Delta Y^R_{i+1}|^2) + 3(\Delta_i + \frac{1}{\gamma_i})\Delta_i\mathbb{E}_i(\Delta f^2_i). \tag{2.3.3}$$

Proof. Preliminary estimates for ΔZ^R_i. From (2.3.2) we have

$$\Delta_i \Delta Z^R_i = \mathbb{E}_i([\Delta Y^R_{i+1} - \mathbb{E}_i(\Delta Y^R_{i+1})][\Delta W^\top_i]_w).$$

By the Cauchy-Schwarz inequality, note that

$$|\mathbb{E}_i([\Delta Y^R_{i+1} - \mathbb{E}_i(\Delta Y^R_{i+1})][\Delta W^\top_i]_w)|^2 \le q\ \Delta_i\Big(\mathbb{E}_i[(\Delta Y^R_{i+1})^2] - (\mathbb{E}_i\Delta Y^R_{i+1})^2\Big)$$

uniformly in R, whence

$$\Delta_i|\Delta Z^R_i|^2 \le q\Big(\mathbb{E}_i[(\Delta Y^R_{i+1})^2] - (\mathbb{E}_i\Delta Y^R_{i+1})^2\Big). \tag{2.3.4}$$

Estimates for ΔY^R_i. We have

$$\Delta Y^R_i = \mathbb{E}_i\Delta Y^R_{i+1} + \Delta_i\mathbb{E}_i[\Delta f_i] + \Delta_i\mathbb{E}_i[\bar{f}_i(Y^R_{i+1}, Z^R_i) - \bar{f}_i(\bar{Y}^R_{i+1}, \bar{Z}^R_i)].$$

Combining the Young inequality $(a+b)^2 \le (1+\gamma_i\Delta_i)a^2 + (1+\frac{1}{\gamma_i\Delta_i})b^2$ and the Lipschitz property of $\bar{f}_i$ and (2.3.4), we deduce

$$
\begin{aligned}
(\Delta Y_i^R)^2 &\leq (1+\gamma_i\Delta_i)(\mathbb{E}_i\Delta Y_{i+1}^R)^2 && (2.3.5)\\
&\quad + 3(\Delta_i + \frac{1}{\gamma_i})\Delta_i\left[\mathbb{E}_i[\Delta f_i^2] + L_{\bar{f}_i}^2\mathbb{E}_i[(\Delta Y_{i+1}^R)^2] + L_{\bar{f}_i}^2|\Delta Z_i^R|^2\right]\\
&\leq \left(1+\gamma_i\Delta_i - 3qL_{\bar{f}_i}^2(\Delta_i+\frac{1}{\gamma_i})\right)(\mathbb{E}_i\Delta Y_{i+1}^R)^2 + 3(\Delta_i+\frac{1}{\gamma_i})\Delta_i\mathbb{E}_i[\Delta f_i^2]\\
&\quad + \left[3(\Delta_i+\frac{1}{\gamma_i})\Delta_i L_{\bar{f}_i}^2 + 3qL_{\bar{f}_i}^2(\Delta_i+\frac{1}{\gamma_i})\right]\mathbb{E}_i[(\Delta Y_{i+1}^R)^2]. && (2.3.6)
\end{aligned}
$$

Under our assumptions on γ_i, we have $\gamma_i \geq 3qL_{\bar{f}_i}^2$, which ensures $1+\gamma_i\Delta_i - 3qL_{\bar{f}_i}^2(\Delta_i+\frac{1}{\gamma_i}) \geq 0$ for any Δ_i. This allows us to combine terms of $(\mathbb{E}_i\Delta Y_{i+1}^R)^2$ and $\mathbb{E}_i[(\Delta Y_{i+1}^R)^2]$ using Jensen's inequality in (2.3.6):

$$
\begin{aligned}
(\Delta Y_i^R)^2 &\leq \left(1+\gamma_i\Delta_i - 3qL_{\bar{f}_i}^2(\Delta_i+\frac{1}{\gamma_i})\right)\mathbb{E}_i[(\Delta Y_{i+1}^R)^2] + 3(\Delta_i+\frac{1}{\gamma_i})\Delta_i\mathbb{E}_i(\Delta f_i^2)\\
&\quad + \left[3(\Delta_i+\frac{1}{\gamma_i})\Delta_i L_{\bar{f}_i}^2 + 3qL_{\bar{f}_i}^2(\Delta_i+\frac{1}{\gamma_i})\right]\mathbb{E}_i[(\Delta Y_{i+1}^R)^2]\\
&= \left(1+\gamma_i\Delta_i + 3(\Delta_i+\frac{1}{\gamma_i})\Delta_i L_{\bar{f}_i}^2\right)\mathbb{E}_i[(\Delta Y_{i+1}^R)^2] + 3(\Delta_i+\frac{1}{\gamma_i})\Delta_i\mathbb{E}_i(\Delta f_i^2),
\end{aligned}
$$

which proves (2.3.3) since $3(\Delta_i+\frac{1}{\gamma_i})\Delta_i L_{\bar{f}_i}^2 \leq \frac{\Delta_i}{2}$. □

The following Proposition will be used extensively in the statistical analysis:

Proposition 2.3.3 (Global pointwise estimates). *Assume that, for each i, $\bar{f}_i$ is Lipschitz w.r.t. y and z with Lipschitz constant $L_{\bar{f}_i} \in \mathbb{R}^+$. Then, for any $R \in [0,+\infty]$, and any time grid π and $\gamma \in (0,+\infty)^N$ satisfying $6q(\Delta_k+\frac{1}{\gamma_k})L_{\bar{f}_k}^2 \leq 1$ for all $k \leq N-1$, we have for $0 \leq i \leq N$*

$$
\begin{aligned}
&|\Delta Y_i^R|^2\Gamma_i + \sum_{k=i}^{N-1}\Delta_k\mathbb{E}_i(|\Delta Z_k^R|^2)\Gamma_k\\
&\qquad \leq C_{2.3.7}\Big(\Gamma_N\mathbb{E}_i(\Delta\xi^2) + 3\sum_{k=i}^{N-1}(\frac{1}{\gamma_k}+\Delta_k)\Delta_k\mathbb{E}_i(\Delta f_k^2)\Gamma_k\Big), \qquad (2.3.7)
\end{aligned}
$$

where $\Gamma_i := \prod_{k=0}^{i-1}(1+\gamma_k\Delta_k)$ and $C_{2.3.7} := 2q + (1+T)e^{T/2}$.

Note that, whenever necessary, the above pointwise estimates can be easily turned into uniform $\mathbf{L}_2$-estimates:

$$
\begin{aligned}
&\sup_{i\leq k\leq N}\mathbb{E}(|\Delta Y_k^R|^2)\Gamma_k + \sum_{k=i}^{N-1}\Delta_k\mathbb{E}(|\Delta Z_k^R|^2)\Gamma_k\\
&\qquad \leq C_{2.3.7}\Big(\Gamma_N\mathbb{E}(\Delta\xi^2) + 3\sum_{k=i}^{N-1}(\frac{1}{\gamma_k}+\Delta_k)\Delta_k\mathbb{E}(\Delta f_k^2)\Gamma_k\Big).
\end{aligned}
$$

Proof. Starting at (2.3.3), multiply both sides by

$$\lambda_i := (1+(\gamma_{i-1}+\frac{1}{2})\Delta_{i-1})\lambda_{i-1}, \quad \lambda_0 := 1,$$

sum between $k = i$ to $k = N-1$, and take conditional expectations $\mathbb{E}_i$ to deduce:

$$(\Delta Y_i^R)^2 \lambda_i \leq \lambda_N \mathbb{E}_i(\Delta\xi^2) + 3\sum_{k=i}^{N-1}(\frac{1}{\gamma_k} + \Delta_k)\Delta_k \mathbb{E}_i(\Delta f_k^2)\lambda_k. \tag{2.3.8}$$

From the simple inequality $\Gamma_i \leq \lambda_i = e^{\sum_{k=0}^{i}\ln(1+(\gamma_k+\frac{1}{2})\Delta_k)} \leq e^{T/2}\Gamma_i$, we get for $0 \leq i \leq N$

$$(\Delta Y_i^R)^2 \Gamma_i \leq e^{T/2}\Gamma_N \mathbb{E}_i(\Delta\xi^2) + 3e^{T/2}\sum_{k=i}^{N-1}(\frac{1}{\gamma_k} + \Delta_k)\Delta_k \mathbb{E}_i(\Delta f_k^2)\Gamma_k. \tag{2.3.9}$$

Final estimates for ΔZ_i. From (2.3.4), we have

$$\begin{aligned}
\sum_{k=i}^{N-1}\Delta_k \mathbb{E}_i[|\Delta Z_k^R|^2]\Gamma_k &\leq \sum_{k=i}^{N-1} q\Gamma_{k+1}\big(\mathbb{E}_i[(\Delta Y_{k+1}^R)^2] - \mathbb{E}_i[(\mathbb{E}_k \Delta Y_{k+1}^R)^2]\big) \\
&\leq q\Gamma_N \mathbb{E}_i(\Delta\xi^2) + \sum_{k=i+1}^{N-1} q\Gamma_k\Big(\mathbb{E}_i[(\Delta Y_k^R)^2] - (1+\gamma_k\Delta_k)\mathbb{E}_i[(\mathbb{E}_k \Delta Y_{k+1}^R)^2]\Big).
\end{aligned}$$

From (2.3.5), we have

$$\begin{aligned}
&\mathbb{E}_i[(\Delta Y_k^R)^2] - (1+\gamma_k\Delta_k)\mathbb{E}_i[(\mathbb{E}_k \Delta Y_{k+1}^R)^2] \\
&\qquad \leq 3(\frac{1}{\gamma_k} + \Delta_k)\Delta_k\Big[\mathbb{E}_i(\Delta f_k^2) + L_{\bar{f}_k}^2 \mathbb{E}_i[(\Delta Y_{k+1}^R)^2] + L_{\bar{f}_k}^2 \mathbb{E}_i[|\Delta Z_k^R|^2]\Big].
\end{aligned}$$

Plugging this inequality into that above yields

$$\begin{aligned}
&\sum_{k=i}^{N-1}\Delta_k \mathbb{E}_i[|\Delta Z_k^R|^2]\Gamma_k \\
&\leq q\Gamma_N \mathbb{E}_i(\Delta\xi^2) + 3\sum_{k=i+1}^{N-1} q(\frac{1}{\gamma_k} + \Delta_k)\Delta_k L_{\bar{f}_k}^2 \mathbb{E}_i(|\Delta Z_k^R|^2)\Gamma_k \\
&+ 3\sum_{k=i+1}^{N-1} q(\frac{1}{\gamma_k} + \Delta_k)\Delta_k \mathbb{E}_i(\Delta f_k^2)\Gamma_k + 3\sum_{k=i+1}^{N-1} q(\frac{1}{\gamma_k} + \Delta_k)\Delta_k L_{\bar{f}_k}^2 \mathbb{E}_i[(\Delta Y_{k+1}^R)^2]\Gamma_k.
\end{aligned}$$

For γ_k and Δ_k as in the Proposition statement, we have $3q(\frac{1}{\gamma_k} + \Delta_k)L_{\bar{f}_k}^2 \leq \frac{1}{2}$, and thus

$$\begin{aligned}
&\sum_{k=i}^{N-1}\Delta_k \mathbb{E}_i[|\Delta Z_k^R|^2]\Gamma_k \\
&\leq 2q\Gamma_N \mathbb{E}_i(\Delta\xi^2) + 6\sum_{k=i+1}^{N-1} q(\frac{1}{\gamma_k} + \Delta_k)\Delta_k \mathbb{E}_i(\Delta f_k^2)\Gamma_k + \sum_{k=i+1}^{N-1}\Delta_k \mathbb{E}_i[(\Delta Y_{k+1}^R)^2]\Gamma_k \\
&\leq (2q + Te^{T/2})\Gamma_N \mathbb{E}_i(\Delta\xi^2) + (6q + 3Te^{T/2})\sum_{k=i+1}^{N-1}(\frac{1}{\gamma_k} + \Delta_k)\Delta_k \mathbb{E}_i(\Delta f_k^2)\Gamma_k,
\end{aligned}$$

where we have used the estimate (2.3.9) on ΔY in the last inequality. $\square$

2.3.2 Projection errors for the ODP and MDP-based schemes

Projection on a closed convex subspace of $\mathbf{L}_2$.

Definition 2.3.4. *Let $\mathcal{S}$ be a non-empty closed convex subset of $\mathbf{L}_2(\mathcal{F}_T, \mathbb{P})$. Then, to any random variable $U \in \mathbf{L}_2(\mathcal{F}_T, \mathbb{P})$ we can associate $\mathcal{P}(U) \in \mathcal{S}$, the (unique) projection of U on $\mathcal{S}$, which satisfies $\mathbb{E}|U - \mathcal{P}(U)|^2 = \inf_{S\in\mathcal{S}} \mathbb{E}|U - S|^2$. For any $S \in \mathcal{S}$, we have*

$$\mathbb{E}\big((U - \mathcal{P}(U))(S - \mathcal{P}(U)\big) \le 0. \tag{2.3.10}$$

$\mathcal{S}$ can be a finite dimensional vector space, i.e. $\mathcal{S} := \{S = S_0 + \sum_{k=1}^K \alpha_k p_k, (\alpha_k)_{1\le k\le K} \in \mathbb{R}^K\}$ for some $p_k \in \mathbf{L}_2$: this is our choice (with $S_0 = 0$) in Section 2.4. It can also be a convex ball of the form $\mathcal{S} = \{S = S_0 + \sum_{k=1}^K \alpha_k p_k, (\alpha_k)_{1\le k\le K} \in \mathbb{R}^K, \|\alpha\| \le \rho\}$ $(\rho \ge 0)$ where $\|\cdot\|$ is *a norm* in $\mathbb{R}^K$. When $\|\cdot\|$ is the Euclidean norm, we obtain the ridge regression [GVL96, Section 12.1], whereas the ℓ_1-norm leads to the Lasso technique [Tib96], providing sparsity in the coefficients.

The projection operator $\mathcal{P}$ satisfies some simple but important properties:

- If $\mathcal{S}$ consists of $\mathcal{Q}$-measurable random variables, then $\mathcal{P}(U) = \mathcal{P}(\mathbb{E}(U|\mathcal{Q}))$. Indeed, $\mathcal{P}(U)$ is the minimizer over $S \in \mathcal{S}$ of $\mathbb{E}|U - S|^2 = \mathbb{E}|U - \mathbb{E}(U|\mathcal{Q})|^2 + \mathbb{E}|\mathbb{E}(U|\mathcal{Q}) - S|^2$.
- The operator $\mathcal{P}$ is 1-Lipschitz. Indeed, for any U_1, U_2 in $\mathbf{L}_2$, write

$$\begin{aligned}\mathbb{E}|U_1 - U_2|^2 =& \mathbb{E}|\mathcal{P}(U_1) - \mathcal{P}(U_2)|^2 + \mathbb{E}|U_1 - \mathcal{P}(U_1) - (U_2 - \mathcal{P}(U_2))|^2\\ &+ 2\mathbb{E}\big((\mathcal{P}(U_1) - \mathcal{P}(U_2))(U_1 - \mathcal{P}(U_1) - (U_2 - \mathcal{P}(U_2))\big)\\ \ge& \mathbb{E}|\mathcal{P}(U_1) - \mathcal{P}(U_2)|^2 + \mathbb{E}|U_1 - \mathcal{P}(U_1) - (U_2 - \mathcal{P}(U_2))|^2\\ \ge& \mathbb{E}|\mathcal{P}(U_1) - \mathcal{P}(U_2)|^2\end{aligned}$$

  using (2.3.10) with $U = U_i$ and $S = \mathcal{P}(U_j)$, $1 \le i \ne j \le 2$.

Projection operators in the DP equations. In the following of this section, the conditional expectation operators in the DP equations will be replaced by projection operators. That is, for each $i \in \{0, \dots, N-1\}$, we consider $\mathcal{S}_i^Y, \mathcal{S}_i^{Z,1}, \dots, \mathcal{S}_i^{Z,q}$ that are non-empty closed convex subsets of $\mathbf{L}_2(\mathcal{F}_{t_i}, \mathbb{P})$. We let $\mathcal{P}_i^Y$ and $\mathcal{P}_i^{Z,1}, \dots, \mathcal{P}_i^{Z,q}$ be the related projection operators and denote the tensor projection $\mathcal{P}^Z := (\mathcal{P}^{Z,l}, \dots, \mathcal{P}^{Z,q})$. We sum up the above stated properties of the $\mathcal{P}_i$ $(= \mathcal{P}_i^Y$ or $\mathcal{P}_i^Z)$:

Lemma 2.3.5. *Let U and V be in $\mathbf{L}^2(\mathcal{F}_T, \mathbb{P})$. Then, we have*

(a) $\mathcal{P}_i(U) = \mathcal{P}_i(\mathbb{E}_i(U))$,

(b) $|\mathcal{P}_i(U) - \mathcal{P}_i(V)|_{\mathbf{L}_2} \le |U - V|_{\mathbf{L}_2}$.

MDP scheme with projection. Using the above projection operators in the discrete BSDE (2.3.1), we obtain the following approximation scheme:

$$\begin{cases} \hat{Y}_i^R &= \mathcal{P}_i^Y\Big(\xi + \sum_{k=i}^{N-1} f_k(\hat{Y}_{k+1}^R, \hat{Z}_k^R)\Delta_k\Big),\\ \Delta_i \hat{Z}_{l,i}^R &= \mathcal{P}_i^{Z,l}\Big([\xi + \sum_{k=i+1}^{N-1} f_k(\hat{Y}_{k+1}^R, \hat{Z}_k^R)\Delta_k][\Delta W_{l,i}]_w\Big),\end{cases} \tag{2.3.11}$$

for $R \in [0, +\infty]$. The following theorem estimates the error between (Y^R, Z^R) and $(\hat{Y}^R, \hat{Z}^R)$.

Theorem 2.3.6. *Assume* $(\mathbf{A}_\xi)$*and* $(\mathbf{A_F}\text{-i-ii})$. *For a given* $\gamma \in [0, +\infty)^N$ *and* $\Gamma_i := \prod_{k=0}^{i-1}(1 + \gamma_k \Delta_k)$, *we define the weighted time-average of error projections on* Y^R *and* Z^R *as follows:*

$$\mathcal{E}_i^{\mathcal{P},Y}(\gamma) = \sum_{k=i}^{N-1} \Delta_k \mathbb{E}(|Y_k^R - \mathcal{P}_k^Y(Y_k^R)|^2)\Gamma_k,$$
$$\mathcal{E}_i^{\mathcal{P},Z}(\gamma) = \sum_{k=i}^{N-1} \Delta_k \mathbb{E}(|Z_k^R - \mathcal{P}_k^Z(Z_k^R)|^2)\Gamma_k.$$

For any $R \in [0, +\infty]$, *any* π *and any* $\gamma \in (0, +\infty)^N$ *such that* $24C_{2.3.7}(1+T)(1 \vee R_\pi)(\frac{1}{\gamma_k} + \Delta_k)\frac{L_f^2}{(T-t_k)^{1-\theta_L}} \leq 1$ *for any* $k < N$, *we have for any* $0 \leq i \leq N-1$

$$\mathbb{E}(|Y_i^R - \hat{Y}_i^R|^2)\Gamma_i \leq 2\mathbb{E}(|Y_i^R - \mathcal{P}_i^Y(Y_i^R)|^2)\Gamma_i + 2\mathcal{E}_{i+1}^{\mathcal{P},Y}(\gamma) + 2\mathcal{E}_i^{\mathcal{P},Z}(\gamma), \tag{2.3.12}$$
$$\sum_{k=i}^{N-1} \Delta_k \mathbb{E}(|Z_k^R - \hat{Z}_k^R|^2)\Gamma_k \leq 4\mathcal{E}_{i+1}^{\mathcal{P},Y}(\gamma) + 4\mathcal{E}_i^{\mathcal{P},Z}(\gamma). \tag{2.3.13}$$

The choice $\gamma_k = 48C_{2.3.7}(1+T)(1 \vee R_\pi)\frac{L_f^2}{(T-t_k)^{1-\theta_L}}$ obviously implies $24C_{2.3.7}(1+T)(1 \vee R_\pi)(\frac{1}{\gamma_k} + \Delta_k)\frac{L_f^2}{(T-t_k)^{1-\theta_L}} \leq \frac{1}{2} + 24C_{2.3.7}(1+T)(1 \vee R_\pi)C_\pi L_f^2 \leq 1$ for N large enough (assuming $(\mathbf{A_F}\text{-iii})$) and moreover, we derive the easy bounds

$$\begin{aligned} 1 \leq \Gamma_i &\leq \exp(\sum_{k=0}^{N-1} \gamma_k \Delta_k) \leq \exp(\int_0^T \frac{48C_{2.3.7}(1+T)(1 \vee R_\pi)L_f^2}{(T-t)^{1-\theta_L}} dt) \\ &= \exp(\frac{48C_{2.3.7}(1+T)(1 \vee R_\pi)L_f^2 T^{\theta_L}}{\theta_L}) := C_{2.3.14}, \end{aligned} \tag{2.3.14}$$

which remains bounded π-uniformly as $N \to +\infty$ owing to (2.2.4). As a consequence, we obtain

Corollary 2.3.7. *Assume* $(\mathbf{A}_\xi)$*and* $(\mathbf{A_F})$. *For any* $R \in [0, +\infty]$ *and for any* π *with* N *large enough (such that* $(1 \vee R_\pi)C_\pi L_f^2 \leq \frac{1}{48C_{2.3.7}(1+T)}$*), we have for any* $0 \leq i \leq N-1$

$$\mathbb{E}(|Y_i^R - \hat{Y}_i^R|^2) \leq 2\mathbb{E}(|Y_i^R - \mathcal{P}_i^Y(Y_i^R)|^2) + 2C_{2.3.14}\big[\mathcal{E}_{i+1}^{\mathcal{P},Y}(0) + \mathcal{E}_i^{\mathcal{P},Z}(0)\big], \tag{2.3.15}$$
$$\sum_{k=i}^{N-1} \Delta_k \mathbb{E}(|Z_k^R - \hat{Z}_k^R|^2) \leq 4C_{2.3.14}\big[\mathcal{E}_{i+1}^{\mathcal{P},Y}(0) + \mathcal{E}_i^{\mathcal{P},Z}(0)\big]. \tag{2.3.16}$$

These estimates show how the error due to projections in the MDP scheme is controlled by the time-average of the projection errors on Y and Z. Moreover, we obtain similar estimates to the Bender-Denk scheme (compare [BD07, Theorem 11] with our estimates), but we avoid the Picard iterations.

Proof.(of Theorem 2.3.6). We first prove a weaker result on the global error, that is

$$\begin{aligned}\mathcal{E}_i(\gamma) &:= \sum_{k=i+1}^{N-1} \Delta_k \mathbb{E}(|Y_k^R - \hat{Y}_k^R|^2)\Gamma_k + \sum_{k=i}^{N-1} \Delta_k \mathbb{E}(|Z_k^R - \hat{Z}_k^R|^2)\Gamma_k \\ &\leq 4\mathcal{E}_{i+1}^{\mathcal{P},Y}(\gamma) + 4\mathcal{E}_i^{\mathcal{P},Z}(\gamma).\end{aligned} \tag{2.3.17}$$

We will also make use of the following intermediate process:

$$\begin{cases} \bar{Y}_i^R &= \mathbb{E}_i\left[\xi + \sum_{k=i}^{N-1} f_k(\hat{Y}_{k+1}^R, \hat{Z}_k^R)\Delta_k\right], \\ \Delta_i \bar{Z}_i^R &= \mathbb{E}_i\left[(\xi + \sum_{k=i+1}^{N-1} f_k(\hat{Y}_{k+1}^R, \hat{Z}_k^R)\Delta_k)[\Delta W_i^\top]_w\right]. \end{cases}$$

Observe that, from Lemma 2.3.5(1), one has the useful properties (for $i < N$)

$$\hat{Y}_i^R = \mathcal{P}_i^Y(\bar{Y}_i^R) \qquad \text{and} \qquad \hat{Z}_i^R = \mathcal{P}_i^Z(\bar{Z}_i^R). \tag{2.3.18}$$

Moreover, $(\bar{Y}^R, \bar{Z}^R)$ solves a discrete BSDE with truncated Brownian increments and data $(\xi, \bar{f}_k : (y,z) \mapsto f_k(\hat{Y}_k^R, \hat{Z}_k^R))$; the Lipschitz constant of $\bar{f}_k$ equals zero for all k. Using Cauchy's inequality and Lemma 2.3.5(2), we obtain

$$\begin{aligned}\mathcal{E}_i(\gamma) \leq\ & 2\sum_{k=i+1}^{N-1} \Delta_k \mathbb{E}(|Y_k^R - \mathcal{P}_k^Y(Y_k^R)|^2)\Gamma_k + 2\sum_{k=i}^{N-1} \Delta_k \mathbb{E}(|Z_k^R - \mathcal{P}_k^Z(Z_k^R)|^2)\Gamma_k \\ &+ 2\sum_{k=i+1}^{N-1} \Delta_k \mathbb{E}(|\mathcal{P}_k^Y(Y_k^R) - \hat{Y}_k^R|^2)\Gamma_k + 2\sum_{k=i}^{N-1} \Delta_k \mathbb{E}(|\mathcal{P}_k^Z(Z_k^R) - \hat{Z}_k^R|^2)\Gamma_k \\ \leq\ & 2\mathcal{E}_{i+1}^{\mathcal{P},Y}(\gamma) + 2\mathcal{E}_i^{\mathcal{P},Z}(\gamma) + 2\sum_{k=i+1}^{N-1} \Delta_k \mathbb{E}(|Y_k^R - \bar{Y}_k^R|^2)\Gamma_k + 2\sum_{k=i}^{N-1} \Delta_k \mathbb{E}(|Z_k^R - \bar{Z}_k^R|^2)\Gamma_k.\end{aligned}$$

To bound the last two terms in the above inequality, we apply Proposition 2.3.3 on the BSDEs (Y^R, Z^R) and $(\bar{Y}^R, \bar{Z}^R)$ to get

$$\begin{aligned}&2\sum_{k=i+1}^{N-1} \Delta_k \mathbb{E}(|Y_k^R - \bar{Y}_k^R|^2)\Gamma_k + 2\sum_{k=i}^{N-1} \Delta_k \mathbb{E}(|Z_k^R - \bar{Z}_k^R|^2)\Gamma_k \\ &\leq 6C_{2.3.7}\big(1 + \sum_{k=i+1}^{N-1} \Delta_k\big) \sum_{k=i}^{N-1} (\frac{1}{\gamma_k} + \Delta_k)\Delta_k \mathbb{E}\big(|f_k(Y_{k+1}^R, Z_k^R) - f_k(\hat{Y}_{k+1}^R, \hat{Z}_k^R)|^2\big)\Gamma_k \\ &\leq 12C_{2.3.7}(1+T) \sum_{k=i}^{N-1} (\frac{1}{\gamma_k} + \Delta_k)\Delta_k \frac{L_f^2}{(T-t_k)^{1-\theta_L}} \mathbb{E}\big(|Y_{k+1}^R - \hat{Y}_{k+1}^R|^2 + |Z_k^R - \hat{Z}_k^R|^2\big)\Gamma_k \\ &\leq \frac{1}{2}\mathcal{E}_i(\gamma).\end{aligned}$$

The penultimate inequality follows from $\Delta_k \leq R_\pi \Delta_{k+1}$ and the conditions on π and γ in the theorem statement. To sum up, we have obtained $\mathcal{E}_i(\gamma) \leq 2\mathcal{E}_{i+1}^{\mathcal{P},Y}(\gamma) + 2\mathcal{E}_i^{\mathcal{P},Z}(\gamma) + \frac{1}{2}\mathcal{E}_i(\gamma)$, which readily proves (2.3.17). This also implies (2.3.13).

We now prove (2.3.12). Proceeding similarly, we obtain

$$
\begin{aligned}
&\mathbb{E}(|Y_i^R - \hat{Y}_i^R|^2)\Gamma_i \le 2\mathbb{E}(|Y_i^R - \mathcal{P}_i^Y(Y_i^R)|^2)\Gamma_i + 2\mathbb{E}(|Y_i^R - \bar{Y}_i^R|^2)\Gamma_i \\
&\le 2\mathbb{E}(|Y_i - \mathcal{P}_i^Y(Y_i^R)|^2)\Gamma_i \\
&\quad + 12C_{2.3.7}\sum_{k=i}^{N-1}(\frac{1}{\gamma_k} + \Delta_k)\Delta_k \frac{L_f^2}{(T-t_k)^{1-\theta_L}}\mathbb{E}\big(|Y_{k+1}^R - \hat{Y}_{k+1}^R|^2 + |Z_k^R - \hat{Z}_k^R|^2\big)\Gamma_k \\
&\le 2\mathbb{E}(|Y_i - \mathcal{P}_i^Y(Y_i^R)|^2)\Gamma_i + \frac{1}{2}\mathcal{E}_i(\gamma)
\end{aligned}
$$

and the proof is complete using (2.3.17). $\square$

Comparison with the ODP scheme with projection. The ODP equation associated with the MDP equation (2.3.11) is

$$
\begin{cases}
\check{Y}_N^R = \xi, \quad \check{Y}_k^R = \mathcal{P}_k^Y\left(\check{Y}_{k+1}^R + f_k(\check{Y}_{k+1}^R, \check{Z}_k^R)\Delta_k\right), \\
\Delta_k \check{Z}_{l,k}^R = \mathcal{P}_k^{Z,l}\left(\check{Y}_{k+1}^R[\Delta W_{l,k}]_w\right).
\end{cases}
$$

In this case, the ODP and MDP equations do not match up, because projection operators do not in general benefit from a tower law. This means that the error analysis would be fundamentally different, because (2.3.18) would no longer be true and one could not apply the stability result for discrete BSDEs. In fact, for uniform time-grid π ($\Delta_k = \frac{T}{N}$ for all k) one would need to multiply the $\mathcal{E}_{i+1}^{\mathcal{P},Y}(0)$ term in (2.3.15) and (2.3.16) by N for the ODP estimates. This is also observed in the error analysis of [LGW06, Theorem 2].

2.3.3 Application of a priori estimates to almost sure bounds

When the terminal condition is bounded, pointwise bounds on Y^R and Z^R are available. These bounds are used in Section 2.4.

Proposition 2.3.8 (*a.s.* upper bounds)**.** *Assume* $(\mathbf{A}'_\xi\textbf{-i})$ *and* $(\mathbf{A_F})$*. For any* $R \in [0, +\infty]$ *and for any* π *with* N *large enough (such that* $C_\pi L_f^2 \le \frac{1}{12q}$*), the following almost sure bounds on* Y_i^R *and* Z_i^R *apply:*

$$
|Y_i^R| \le C_y := C_{2.3.19}\Big(C_\xi + \frac{T^{\theta_c}}{\sqrt{4q(2\theta_c - \theta_L \wedge \theta_c)}}C_f\Big), \quad |Z_{l,i}^R| \le C_{z,i} := \frac{C_y}{\sqrt{\Delta_i}}, \tag{2.3.19}
$$

for $0 \le i \le N-1$ *and* $C_{2.3.19} = \exp\big(\frac{T}{4} + \frac{6q(1\vee L_f^2)}{\theta_L \wedge \theta_c}(T^{\theta_L} \vee 1)\big)$.

Observe that C_y and $C_{2.3.19}$ are uniform in i and $R \in [0, +\infty]$, and that they remain bounded as L_f and T go to 0 (as we naturally expect).

Proof. We derive the almost sure bounds from the global pointwise estimates in Proposition 2.3.3. To apply the results of this proposition, we take the pair $(0, 0)$ for the first discrete BSDE - the solution associated to the null driver and terminal condition - and (Y^R, Z^R) for the second discrete BSDE, which is given by the DP equation (2.3.1). From (2.3.8), for any Δ_i and γ_i such that $6q(\Delta_i + \frac{1}{\gamma_i})\frac{L_f^2}{(T-t_i)^{1-\theta_L}} \leq 1$, and recalling that $\lambda_i := \prod_{k=0}^{i-1}(1 + (\gamma_k + \frac{1}{2})\Delta_k)$, we have

$$
\begin{aligned}
(\Delta Y_i^R)^2 \leq (\Delta Y_i^R)^2 \lambda_i &\leq \lambda_N \mathbb{E}_i(\xi^2) + 3\sum_{k=i}^{N-1} \frac{(1+\gamma_k\Delta_k)}{\gamma_k}\lambda_k\Delta_k\mathbb{E}_i(f_k^2(0,0)) \\
&\leq \lambda_N \big(C_\xi^2 + 3C_f^2 \sum_{k=0}^{N-1} \frac{\Delta_k}{\gamma_k(T-t_k)^{2(1-\theta_c)}}\big).
\end{aligned}
$$

For N large enough, we have $C_\pi L_f^2 \leq \frac{1}{12q}$; additionally, we set

$$
\gamma_k := 12q\frac{(1\vee T^{-\theta_L})(1\vee L_f^2)}{(T-t_k)^{1-\theta_L}}\big(\frac{T}{T-t_k}\big)^{\theta_L-\theta_L\wedge\theta_c} \geq \frac{12qL_f^2}{(T-t_k)^{1-\theta_L}}, \qquad 0 \leq k < N.
$$

It follows $6q(\Delta_k + \frac{1}{\gamma_k})\frac{L_f^2}{(T-t_k)^{1-\theta_L}} \leq 1$. Easy computations similar to (2.3.14) give

$$
\begin{aligned}
\lambda_N &\leq \exp\big(\frac{T}{2} + 12q(1\vee T^{-\theta_L})(1\vee L_f^2)T^{\theta_L-\theta_L\wedge\theta_c}\int_0^T (T-t)^{\theta_L\wedge\theta_c-1}dt\big) \\
&= \exp\big(\frac{T}{2} + \frac{12q(1\vee L_f^2)}{\theta_L\wedge\theta_c}(T^{\theta_L}\vee 1)\big), \\
\sum_{k=0}^{N-1}\frac{\Delta_k}{\gamma_k(T-t_k)^{2(1-\theta_c)}} &= \sum_{k=0}^{N-1}\frac{\Delta_k(1\wedge T^{\theta_L})T^{\theta_L\wedge\theta_c-\theta_L}}{12q(1\vee L_f^2)(T-t_k)^{(1-2\theta_c+\theta_L\wedge\theta_c)}} \\
&\leq \frac{T^{2\theta_c}}{12q(2\theta_c-\theta_L\wedge\theta_c)}
\end{aligned}
$$

(observing that $2\theta_c - \theta_L\wedge\theta_c \geq \theta_c > 0$). Combining the last three inequalities, we obtain the required upper bounds (2.3.19) on Y^R. The bound on Z_i^R is clear from the Cauchy-Schwartz inequality and the bound on Y_{i+1}^R. □

2.3.4 Impact of the threshold R

In the spirit of [LGW06], we compare the discrete BSDE (2.3.2) (or equivalently (2.3.1)) with $R < +\infty$, to (2.1.1) (or equivalently (2.1.4)). While unessential, we assume for simplicity that the terminal condition is bounded.

Proposition 2.3.9. *Assume* $(\mathbf{A}'_\xi\textbf{-i})$*and* $(\mathbf{A_F})$*. For any* $R \in [0,+\infty]$ *and for any* π *with* N *large enough (such that* $C_\pi L_f^2 \leq \frac{1}{12q}$*), the following almost sure error bounds on* $Y_i - Y_i^R$ *and* $Z_i - Z_i^R$ *hold for any* $0 \leq i < N$*:*

$$
|Y_i - Y_i^R| \leq C_y \exp\big(\frac{T}{8} + \frac{12qL_f^2}{\theta_L}T^{\theta_L}\big)\exp\big(-\frac{1}{4}R^2\big)\sqrt{N},
$$

$$
\big(\sum_{k=i}^{N-1}\mathbb{E}_i|Z_k - Z_k^R|^2\Delta_k\big)^{\frac{1}{2}} \leq C_y \exp\big(\frac{12qL_f^2}{\theta_L}T^{\theta_L}\big)\big(8q + T\exp(\frac{T}{4})\big)^{\frac{1}{2}}\exp\big(-\frac{1}{4}R^2\big)\sqrt{N}.
$$

For the proof, see Section 2.5.2. Consequently, taking R in a logarithmic scale w.r.t. N is sufficient to make the threshold error negligible: for instance, taking $R = \sqrt{4(p+\frac{1}{2})\log(N+1)}$ (for $p \geq 0$) gives an error of magnitude $O(N^{-p})$.

2.3.5 Application of a priori bounds and Markov assumptions to additional smoothness properties

In this subsection, the Markovian assumptions $(\mathbf{A_X})$, $(\mathbf{A'_\xi})$and $(\mathbf{A'_F})$are in force. We demonstrate how additional smoothness conditions of the terminal condition Φ and the driver f_k strongly improve the smoothness properties of y_k^R and z_k^R. Increased smoothness is essential in the complexity analysis of numerical algorithms, as will be demonstrated in Section 2.4.3.

Lemma 2.3.10. *Assume $x \mapsto \Phi(x)$ is uniformly Lipschitz continuous with Lipschitz constant L_Φ, and that the Lipschitz property of the driver f_k is extended to the x component; i.e.,*

$$|f_k(x,y,z) - f_k(x',y',z')| \leq \frac{L_f}{(T-t_k)^{(1-\theta_L)/2}}(|x-x'| + |y-y'| + |z-z'|).$$

Furthermore, assume that, for any $x_1, x_2 \in \mathbb{R}^d$ and $0 \leq k \leq N-1$, two Markov chains $(X_i^{k,x_1})_{k\leq i\leq N}$ and $(X_i^{k,x_2})_{k\leq i\leq N}$ started at time t_k with values x_i $(i=1,2)$ with the same transition probabilities as $(X_k)_k$ enjoy the property $\mathbb{E}[|X_i^{k,x_1} - X_i^{k,x_2}|^2] \leq C_X|x_1 - x_2|^2$ for some constant C_X.

Then, if $12qC_\pi L_f^2 \leq 1$ for all k, $x \mapsto y_k^R(x)$ is Lipschitz continuous uniformly in k and N.

Proof. Let (Y^1, Z^1) and (Y^2, Z^2) be discrete BSDEs from k to N with data $(\Phi(X_N^{k,x_j}), f_i(X_i^{k,x_j}, y, z))$ $(j = 1, 2$ resp.). Then $y_k^R(x_j) = Y_k^j$ and $z_k^R(x_j) = Z_k^j$ $(j = 1, 2$ resp.) hold almost surely. We use the result of (2.3.9), together with the choice $\gamma_k = 12qL_f^2(T-t_k)^{-(1-\theta_L)}$, to obtain

$$\begin{aligned}|y_k^R(x_1) - y_k^R(x_2)|^2 &\leq e^{T/2}C_\Gamma \mathbb{E}[|\Phi(X_N^{k,x_1}) - \Phi(X_N^{k,x_2})|^2] \\ &\quad + 3e^{T/2}C_\Gamma \sum_{i=k}^{N-1}(\Delta_i + \frac{1}{\gamma_i})\mathbb{E}[|f_i(X_i^{k,x_1}, Y_{i+1}^1, Z_i^1) - f_i(, X_i^{k,x_2}, Y_{i+1}^1, Z_i^1)|^2]\Delta_i \\ &\leq e^{T/2}(\frac{T}{2q} + L_\Phi^2)C_X C_\Gamma |x_1 - x_2|^2 \qquad (2.3.20)\end{aligned}$$

where $C_\Gamma := e^{12qL_f^2 T^{\theta_L}/\theta_L}$ comes from the choice of γ_k; see (2.3.14) for details. □

We now show that, under assumptions of Lipschitz continuous data, the a.s. bound of Z does not suffer from the inverse dependency on the time increments.

Corollary 2.3.11. *Under the assumptions of Lemma 2.3.10 and additionally that for all $x \in \mathbb{R}^d$, the Markov chain $(X_i^{k,x})_{i\geq k}$ started at t_k with value x enjoys the property that $\mathbb{E}[|X_{k+1}^{k,x} - x|^2] \leq C_X\Delta_k$, then the function $x \mapsto z_k^R(x)$ is uniformly bounded, with a bound independent of the time increments.*

Proof. The result follows directly from Lemma 2.3.10:

$$|z_k^R(x)|^2 = \frac{1}{\Delta_k^2}|\mathbb{E}[[\Delta W_k]_w(y_{k+1}^R(X_{k+1}^{k,x}) - y_{k+1}^R(x))]|^2 \leq e^{T/2}(\frac{T}{2} + qL_\Phi^2)C_X^2 C_\Gamma$$

where we have used the fact that y_{k+1}^R is deterministic in the first equality, and the Cauchy-Schwartz inequality combined with (2.3.20) in the second inequality. □

The additional assumptions made in Lemma 2.3.10 and Corollary 2.3.11 are quite natural: they are satisfied by, for example, the Euler scheme for a jump-diffusion with bounded coefficients.

Extensions to Lemma 2.3.10 and Corollary 2.3.11 to higher derivatives will be carried out in future work.

2.4 Empirical regression scheme: a first MDP algorithm

In this section, we approximate the projection scheme (2.3.11) using least-squares regression on simulated data; we call this the LSMDP scheme. The details of the algorithm are made explicit in Section 2.4.1, and a full error analysis is undertaken in Section 2.4.2. Then the algorithm complexity is discussed in Section 2.4.3. Finally, Sections 2.4.4 and 2.4.5 are devoted to the proof of our main results. For this algorithm, we only a single set of independent paths, while in the next section, several independent sets of independent paths are used.

2.4.1 Notation and algorithm

Markovian framework. In what follows, we always assume $(\mathbf{A_X})$, $(\mathbf{A}'_\xi)$and $(\mathbf{A}'_\mathbf{F})$. This allows the representation

$$(Y_i^R, Z_i^R) := (y_i^R(X_i), z_i^R(X_i)) \tag{2.4.1}$$

for measurable, deterministic functions $y_i^R(\cdot)$ and $z_i^R(\cdot)$.

Samples. Let $\{(X_k^m)_{k\geq 0}\}_{m=1,\dots,M}$ denote M independent paths of the Markov chain, and $\{(\Delta W_k^m)_{k\geq 0}\}_{m=1,\dots,M}$ the independent increments of the Brownian Motion from which the Markov chain is generated. We denote the samples of the Markov chain at time k by $X_k^{1:M} := \{X_k^m\}_{m=1,\dots,M}$. For function $\psi : \mathbb{R}^d \to \mathbb{R}$, we define the empirical norm with respect to the sample $X_k^{1:M}$ by

$$\|\psi\|_{k,M} := \big(\frac{1}{M}\sum_{m=1}^{M} |\psi(X_k^m)|^2\big)^{1/2}.$$

Basis functions. For each $l = 0, \dots, q$ and $k = 0, \dots, N-1$, we are given a finite number of deterministic *basis* functions $(p_{l,k}^i(.))_{1\leq i\leq K_{l,k}}$, where $p_{l,k}^i(.) : \mathbb{R}^d \to \mathbb{R}$ satisfies $\mathbb{E}[|p_{l,k}^i(X_k)|^2] < +\infty$. We write the functions as a column vector $p_{l,k}(.) = (p_{l,k}^1(.), \dots)^\top$, where $\top$ denotes the transpose operator. Without loss of generality, we assume that $M \geq K_{l,k}$. The random variables $(p_{l,k}^i(X_k))_{1\leq i\leq K_{l,k}}$ span a linear subspace of $\mathbf{L}_2(\mathcal{F}_{t_k}, \mathbb{P})$, which is denoted $\mathcal{S}_k^Y$ if $l = 0$ and $\mathcal{S}_k^{Z,l}$ if $l \neq 0$ in the language of Subsection 2.3.2. The extension of our error estimates to more general closed convex subspaces will be considered in future research.
We write $p_{l,k}^m$ (resp. $p_{l,k}^{i,m}$) to mean $p_{l,k}(X_k^m)$ (resp. $p_{l,k}^i(X_k^m)$).

Least-squares problem. Instead of projections in the $\mathbf{L}_2(\mathbb{P})$-sense as in Subsection 2.3.2, we numerically compute empirical regressions using the samples. Generally speaking for given observation $X_k^{1:M}$ and response $S^{1:M} = (S^m)_{m=1,\dots,M}$, we aim at computing the best approximation of the response in the vector space generated by the basis functions $p_{l,k}$ w.r.t. the norm $\|.\|_{k,M}$: it is defined by $\alpha^\star \cdot p_{l,k}(.)$ where

$$\alpha^\star = \arg \inf_{\alpha\in\mathbb{R}^{K_{l,k}}} \|\alpha \cdot p_{l,k} - S\|_{k,M}^2. \tag{2.4.2}$$

Since colinearities may exist between basis functions, the above coefficient $\alpha^\star$ may be not unique and one must first clarify which solution to take. We take the Singular Value Decomposition (SVD in short) approach by taking the coefficient with minimal Euclidean norm (see 2.5.1 for details). We refer to this choice as the *SVD-optimal* coefficient. We now state basic properties related to least-squares regression with random observation $X_k^{1:M}$ that will be frequently used in this work.

Proposition 2.4.1. *Let $\alpha^\star$ be the SVD-optimal coefficient solving* $\arg\inf_{\alpha\in\mathbb{R}^{K_{l,k}}} \|\alpha\cdot p_{l,k} - S\|_{k,M}^2$*. The following properties are satisfied:*

i) linearity: the mapping $S \mapsto \alpha^\star$ is linear.

ii) contraction property: $\|\alpha^\star \cdot p_{l,k}\|_{k,M} \le \|S\|_{k,M}$.

iii) conditional expectation solution: assume that $(p_{l,k}^m)_{m=1,\dots,M}$ is measurable with respect to the sub-σ-algebra $\mathcal{Q}$. Then the SVD-optimal coefficient associated to the response $\mathbb{E}(S|\mathcal{Q}) = (\mathbb{E}(S^m|\mathcal{Q}))_{m=1,\dots,M}$ is given by $\mathbb{E}(\alpha^\star|\mathcal{Q})$.

The proof is given in 2.5.1.

Soft thresholds for approximate solutions. y_i^R and z_i^R are bounded by C_y and $C_{z,i}$, respectively, provided that N is large enough (see Proposition 2.3.8). We force the approximated solutions to satisfy these bounds: for $(y,z) := (y,(z_1,\cdots,z_q)) \in \mathbb{R}\times\mathbb{R}^q$ we define the soft thresholding

$$[y]_y = -C_y \vee y \wedge C_y, \quad [z_l]_z = -C_{z,i} \vee z_l \wedge C_{z,i}, \quad [z]_z = ([z_1]_z,\cdots,[z_q]_z). \tag{2.4.3}$$

In the notation for the z-threshold, we do not indicate that it depends on i because this is clear from the context.

Coefficients and solution approximations. We set $y_N^{R,M}(\cdot) := \Phi(\cdot)$. For $k < N$, we iteratively define the SVD-optimal coefficients

$$\begin{aligned} \alpha_{0,k}^M := \arg\min_\alpha \frac{1}{M}\sum_{m=1}^M \big(\Phi(X_N^m) \\ + \sum_{i=k}^{N-1} f_i(X_i^m, y_{i+1}^{R,M}(X_{i+1}^m), z_i^{R,M}(X_i^m))\Delta_i - \alpha\cdot p_{0,k}^m\big)^2, \end{aligned} \tag{2.4.4}$$

$$\begin{aligned} \alpha_{l,k}^M := \arg\min_\alpha \frac{1}{M}\sum_{m=1}^M \Big(\frac{[\Delta W_{l,k}^m]_w}{\Delta_k}\big(\Phi(X_N^m) \\ + \sum_{i=k+1}^{N-1} f_i(X_i^m, y_{i+1}^{R,M}(X_{i+1}^m), z_i^{R,M}(X_i^m))\Delta_i\big) - \alpha\cdot p_{l,k}^m\Big)^2. \end{aligned} \tag{2.4.5}$$

Then we define the following functions used to approximate y_k^R and z_k^R respectively

$$y_k^{R,M}(x) := [\alpha_{0,k}^M \cdot p_{0,k}(x)]_y, \qquad z_{l,k}^{R,M}(x) := [\alpha_{l,k}^M\cdot p_{l,k}(x)]_z, \tag{2.4.6}$$

where the thresholds $[\cdot]_y$ and $[\cdot]_z$ are defined in (2.4.3). Thanks to these thresholds, the random function

$$\Psi_k^{R,M}(x_k,\dots,x_N) := \Phi(x_N) + \sum_{i=k}^{N-1} f_i(x_i, y_{i+1}^{R,M}(x_{i+1}), z_i^{R,M}(x_i))\Delta_i \tag{2.4.7}$$

is bounded independently of the samples and of $(x_k,\ldots,x_N)$; this property will be used repeatedly in the subsequent analysis.

Lemma 2.4.2. *Under* $(\mathbf{A_X})$, $(\mathbf{A}'_\xi)$*and* $(\mathbf{A}'_\mathbf{F})$, *we have*

$$\sup_{0\le k\le N}\ \sup_{x_k\in\mathbb{R}^d,\ldots,x_N\in\mathbb{R}^d}|\Psi_k^{R,M}(x_k,\ldots,x_N)|\le C_\Psi$$

where $C_\Psi := C_\xi + L_f C_y T^{\frac{\theta_L}{2}}\big[\frac{2\sqrt{T}}{1+\theta_L}+\frac{\sqrt{q}\sqrt{N}}{\sqrt{\theta_L}}\big]+C_f\frac{T^{\theta_c}}{\theta_c}$.

Proof. From $(\mathbf{A}'_\xi$-**i**$)$, $(\mathbf{A_F}$-**i-ii**$)$, (2.4.6) and (2.4.3), we readily obtain

$$\begin{aligned}|\Psi_k^{R,M}(x_k,\ldots,x_N)| &\le C_\xi+\sum_{i=0}^{N-1}\Big[\frac{L_f}{(T-t_i)^{\frac{1-\theta_L}{2}}}(C_y+\sqrt{q}\frac{C_y}{\sqrt{\Delta_i}})\Delta_i+\frac{C_f}{(T-t_i)^{1-\theta_c}}\Delta_i\Big]\\ &\le C_\xi+L_fC_y\Big[\frac{T^{(1+\theta_L)/2}}{(1+\theta_L)/2}+\sqrt{q}\sqrt{N}\Big(\sum_{i=0}^{N-1}\big(\frac{\sqrt{\Delta_i}}{(T-t_i)^{\frac{1-\theta_L}{2}}}\big)^2\Big)^{1/2}\Big]+C_f\frac{T^{\theta_c}}{\theta_c}\end{aligned}$$

and the announced upper bound follows. □

2.4.2 Error analysis

In contrast to standard regression problems, a major difficulty for the error analysis is related to the non-independence of the random variables $\{\Phi(X_N^m)+\sum_{i=k}^{N-1}f_i(X_i^m,y_{i+1}^{R,M}(X_{i+1}^m),z_i^{R,M}(X_i^m))\Delta_i\}_{m=1,\ldots,M}$ due to the interdependence of the random functions $(y_i^{R,M}(.),z^{R,M}(.))_{i=k,\ldots,N-1}$. To deal with this, we follow the method of [LGW06], which uses methods from statistical learning, but introduce some important adaptations. In particular, we use intermediate processes in order to take advantage of the a priori results for discrete BSDEs, leading to important improvements in the error estimates. The subsection is organized as follows: first, we introduce the tools of statistical learning we require, the intermediate processes, and the local error terms; then we state a global error decomposition in terms of the local error terms in Theorem 2.4.4, which is the corner stone of our error analysis, and bound the local error terms in Theorem 2.4.5.

Ghost sample. In the upcoming error analysis, we employ the method of *symmetrization*, which is standard in statistical learning; see [Pol84] or [GKKW02]. This involves the introduction of paths of the Markov chain which are identically distributed but independent (*ghost*) to the original samples.

Let k be given. For each m, we denote by $(\tilde{X}_i^{k,m})_{i\ge k}$ an independent copy of the Markov chain $(X_i^m)_{i\ge k}$ starting at t_k with value X_k^m ($\tilde{X}_k^{k,m}=X_k^m$). Additionally, we denote by $\Delta\tilde{W}_k^{k,m}$ the ghost Brownian increment used to generate the Markov chain $(\tilde{X}_i^{k,m})_{i\ge k}$: it is independent of and identically distributed to ΔW_k^m. Conditionally on

$$\mathcal{F}_k^M := \sigma(X_i^m,\Delta W_{j-1}^m : i,j\le k, 1\le m\le M),$$

the ghost paths $\{(\tilde{X}_i^{k,m})_{i\ge k},\Delta\tilde{W}_k^{k,m} : 1\le m\le M\}$ are independent. Furthermore, we write $\mathbb{E}_k^M$ ($\mathbb{P}_k^M$) for the conditional expectation (probability) with respect to $\mathcal{F}_k^M$.

Extra coefficients. To analyze the convergence, we make use of coefficients calculated using the ghost paths:

$$\begin{aligned}\tilde{\alpha}_{0,k}^{M} := \arg\min_{\alpha} \frac{1}{M} \sum_{m=1}^{M} \big(\Phi(\tilde{X}_N^{k,m}) \\ + \sum_{i=k}^{N-1} f_i(\tilde{X}_i^{k,m}, y_{i+1}^{R,M}(\tilde{X}_{i+1}^{k,m}), z_i^{R,M}(\tilde{X}_i^{k,m}))\Delta_i - \alpha \cdot p_{0,k}^m\big)^2,\end{aligned} \tag{2.4.8}$$

$$\begin{aligned}\tilde{\alpha}_{l,k}^{M} := \arg\min_{\alpha} \frac{1}{M} \sum_{m=1}^{M} \big(\frac{[\Delta \tilde{W}_{l,k}^{k,m}]_w}{\Delta_k} \big(\Phi(\tilde{X}_N^{k,m}) \\ + \sum_{i=k+1}^{N-1} f_i(\tilde{X}_i^{k,m}, y_{i+1}^{R,M}(\tilde{X}_{i+1}^{k,m}), z_i^{R,M}(\tilde{X}_i^{k,m}))\Delta_i\big) - \alpha \cdot p_{l,k}^m\big)^2.\end{aligned} \tag{2.4.9}$$

In addition, we need the following coefficients, also calculated with the ghost paths but with the functions y^R and z^R, from the Markov representation (2.4.1) of (Y^R, Z^R), in the place of $y^{R,M}$ and $z^{R,M}$:

$$\begin{aligned}\tilde{\beta}_{0,k}^{M} := \arg\min_{\alpha} \frac{1}{M} \sum_{m=1}^{M} \big(\Phi(\tilde{X}_N^{k,m}) \\ + \sum_{i=k}^{N-1} f_i(\tilde{X}_i^{k,m}, y_{i+1}^{R}(\tilde{X}_{i+1}^{k,m}), z_i^{R}(\tilde{X}_i^{k,m}))\Delta_i - \alpha \cdot p_{0,k}^m\big)^2,\end{aligned} \tag{2.4.10}$$

$$\begin{aligned}\tilde{\beta}_{l,k}^{M} := \arg\min_{\alpha} \frac{1}{M} \sum_{m=1}^{M} \big(\frac{[\Delta \tilde{W}_{l,k}^{k,m}]_w}{\Delta_k} \big(\Phi(\tilde{X}_N^{k,m}) \\ + \sum_{i=k+1}^{N-1} f_i(\tilde{X}_i^{k,m}, y_{i+1}^{R}(\tilde{X}_{i+1}^{k,m}), z_i^{R}(\tilde{X}_i^{k,m}))\Delta_i\big) - \alpha \cdot p_{l,k}^m\big)^2.\end{aligned} \tag{2.4.11}$$

Intermediate processes. Let $(X_i^{k,x})_{k \le i \le N}$ be a Markov chain starting at t_i with value x with the same transition probabilities as X. We generate the following intermediate (sample dependent) functions:

$$\bar{y}_k^{R,M}(x) := \int \Psi_k^{R,M}(x, x_{k+1}, \ldots, x_N)\mu^x(\mathrm{d}x_{k+1}, \ldots, \mathrm{d}x_N)\ , \tag{2.4.12}$$

$$\Delta_k \bar{z}_k^{R,M}(x) := \int [w]_w^\top \Psi_{k+1}^{R,M}(x_{k+1}, \ldots, x_N)\mu^{x,W}(\mathrm{d}w, \mathrm{d}x_{k+1}, \ldots, \mathrm{d}x_N) \tag{2.4.13}$$

where μ^x is the law of $(X_{k+1}^{k,x}, \ldots, X_N^{k,x})$ and $\mu^{x,W}$ the law of $(\Delta W_k, X_{k+1}^{k,x}, \ldots, X_N^{k,x})$. Using Lemma 2.4.2, we directly derive the upper bounds

$$|\bar{y}_k^{R,M}(x)| \le C_\Psi, \qquad |\bar{z}_{l,k}^{R,M}(x)| \le \frac{C_\Psi}{\sqrt{\Delta_k}}. \tag{2.4.14}$$

Lemma 2.4.3. *With the current notation and assumptions, for all m we have*

$$\bar{y}_k^{R,M}(X_k^m) = \mathbb{E}_N^M\big[\Phi(\tilde{X}_N^{k,m}) + \sum_{i=k}^{N-1} f_i(\tilde{X}_i^{k,m}, y_{i+1}^{R,M}(\tilde{X}_{i+1}^{k,m}), z_i^{R,M}(\tilde{X}_i^{k,m}))\Delta_i\big],$$

$$\Delta_k \bar{z}_{l,k}^{R,M}(X_k^m) = \mathbb{E}_N^M\big[[\Delta\tilde{W}_{l,k}^m]_w\big(\Phi(\tilde{X}_N^{k,m}) + \sum_{i=k+1}^{N-1} f_i(\tilde{X}_i^{k,m}, y_{i+1}^{R,M}(\tilde{X}_{i+1}^{k,m}), z_i^{R,M}(\tilde{X}_i^{k,m}))\Delta_i\big)\big].$$

Proof. We start with a standard result. Let $\mathcal{G}$ and $\mathcal{H}$ be sub-σ-algebras of $\mathcal{F}$, such that $\mathcal{G} \perp\!\!\!\perp \mathcal{H}$. Let $F : \Omega \times \mathbb{R}^d \to \mathbb{R}^d$ be bounded and $\mathcal{G}\otimes\mathcal{B}(\mathbb{R}^d)$-measurable, and $U : \Omega \to \mathbb{R}^d$ be $\mathcal{H}$-measurable. Then, by the Monotone Class Theorem for functions, $\mathbb{E}[F(U)|\mathcal{H}] = j(U)$ where $j(h) = \mathbb{E}[F(h)]$ for all $h \in \mathbb{R}^d$.
In order to apply the above result, we require some standard results about the ghost path $(\tilde{X}, \Delta\tilde{W})$. Let k be fixed. Since $\tilde{X}$ is a Markov chain, then for all $i > k$ there is a mapping $V_i : \Omega \times \mathbb{R}^d \to \mathbb{R}^d$ measurable with respect to $\mathcal{G}_i \otimes \mathcal{B}(\mathbb{R}^d)$ such that $\tilde{X}_i^{x,k} = V_i(x)$, where the filtration $(\mathcal{G}_i)_{k<i\le N}$ is independent of $\mathcal{F}_N^M$ and $\Delta\tilde{W}_k^k$ is $\mathcal{G}_{k+1}$-measurable.
Now, by defining

$$\begin{cases} F_1(x) & := \Psi_k^{R,M}(x, V_{k+1}(x), \ldots, V_N(x)), \\ F_2(x) & := [\Delta\tilde{W}_k^{k,m}]_w \Psi_{k+1}^{R,M}(V_{k+1}(x), V_{k+2}(x), \ldots, V_N(x)), \end{cases}$$

the result of the previous paragraph can be applied, because F_1 and F_2 are $\mathcal{G}_N\otimes\mathcal{B}(\mathbb{R}^d)$-measurable, hence the representations for $\bar{y}_k^{R,M}(X_k^m)$ and $\bar{z}_k^{R,M}(X_k^m)$. $\square$

Local error terms. For given *accuracy parameters* $\varepsilon_{\cdot,A}, \varepsilon_{\cdot,B}, \varepsilon_{\cdot,C} \in (0,+\infty)^{2N}$, we define the events:

$$A_k^{Y,M} := \big\{\|(\alpha_{0,k}^M - \tilde{\alpha}_{0,k}^M)\cdot p_{0,k}\|_{k,M}^2 > \varepsilon_{k,A}^Y\big\}, \tag{2.4.15}$$
$$A_k^{Z,M} := \big\{\exists l \in \{1,\ldots,q\} \text{ s.t. } \|(\alpha_{l,k}^M - \tilde{\alpha}_{l,k}^M)\cdot p_{l,k}\|_{k,M}^2 > \varepsilon_{k,A}^Z\big\},$$
$$B_k^{Y,M} := \big\{\varepsilon_{k,B}^Y + 2\mathbb{E}_N^M[(\bar{y}_k^{R,M}(X_k) - y_k^R(X_k))^2] < \|\bar{y}_k^{R,M} - y_k^R\|_{k,M}^2\big\}, \tag{2.4.16}$$
$$B_k^{Z,M} := \big\{\exists l \in \{1,\ldots,q\} \text{ s.t. } \varepsilon_{k,B}^Z + 2\mathbb{E}_N^M[(\bar{z}_{l,k}^{R,M}(X_k) - z_{l,k}^R(X_k))^2] < \|\bar{z}_{l,k}^{R,M} - z_{l,k}^R\|_{k,M}^2\big\},$$
$$C_k^{Y,M} := \big\{\varepsilon_{k,C}^Y + 2\|y_k^{R,M} - y_k^R\|_{k,M}^2 < \mathbb{E}_N^M[(y_k^{R,M}(X_k) - y_k^R(X_k))^2]\big\}, \tag{2.4.17}$$
$$C_k^{Z,M} := \big\{\exists l \in \{1,\ldots,q\} \text{ s.t. } \varepsilon_{k,C}^Z + 2\|z_{l,k}^{R,M} - z_{l,k}^R\|_{k,M}^2 < \mathbb{E}_N^M[(z_{l,k}^{R,M}(X_k) - z_{l,k}^R(X_k))^2]\big\}.$$

These six events are *large deviation events*. In Theorem 2.4.5, we show that their probabilities are exponentially small under appropriate choice of the accuracy parameters; the exponent ought to depend on the complexity of the class of functions spanned by $p_{l,k}$, on Δ_k, C_y, C_Ψ, R and M. We also consider

$$T_{1,k}^{Y,M} := \mathbb{E}\big(\inf_{\alpha} \|\alpha \cdot p_{0,k} - y_k^R\|_{k,M}^2\big), \tag{2.4.18}$$

$$T_{1,k}^{Z,M} := \sum_{l=1}^{q} \mathbb{E}\big(\inf_{\alpha} \|\alpha \cdot p_{l,k} - z_{l,k}^R\|_{k,M}^2\big), \tag{2.4.19}$$

$$T_{2,k}^{Y,M} := \mathbb{E}\big\|(\tilde{\alpha}_{0,k}^M - \mathbb{E}_N^M[\tilde{\alpha}_{0,k}^M]) \cdot p_{0,k}\big\|_{k,M}^2, \tag{2.4.20}$$

$$T_{2,k}^{Z,M} := \sum_{l=1}^{q} \mathbb{E}\big\|(\tilde{\alpha}_{l,k}^M - \mathbb{E}_N^M[\tilde{\alpha}_{l,k}^M]) \cdot p_{l,k}\big\|_{k,M}^2. \tag{2.4.21}$$

Equations (2.4.18) - (2.4.21) have standard interpretation in regression theory: the two first terms are square bias terms (best approximation error of the basis functions), while the last two are variance terms (statistical errors).

Error decomposition. We now state the main results of the global error analysis. Similarly to Corollary 2.3.7, these global errors read as time-average of local errors.

Theorem 2.4.4 (Error for the LSMDP scheme)**.** *Assume* $(\mathbf{A_X})$, $(\mathbf{A}'_\xi)$ *and* $(\mathbf{A}'_\mathbf{F})$*. Define*

$$\begin{aligned}
\mathcal{E}_k^Y(R,M) := {}& T_{1,k}^{Y,M} + 3T_{2,k}^{Y,M} + 3\varepsilon_{k,A}^Y + 12C_\Psi^2\mathbb{P}(A_k^{Y,M}) \\
&+ 3\varepsilon_{k,B}^Y + 6(C_y^2 + C_\Psi^2)\mathbb{P}(B_k^{Y,M}) + \frac{1}{4}\varepsilon_{k,C}^Y R_\pi + C_y^2 R_\pi \mathbb{P}(C_k^{Y,M}), \\
\mathcal{E}_k^Z(R,M) := {}& T_{1,k}^{Z,M} + 3T_{2,k}^{Z,M} + 3q\varepsilon_{k,A}^Z + 12\frac{C_\Psi^2 R^2}{\Delta_k}\mathbb{P}(A_k^{Z,M}) \\
&+ 3q\varepsilon_{k,B}^Z + 6q\frac{(C_y^2 + C_\Psi^2)}{\Delta_k}\mathbb{P}(B_k^{Z,M}) + \frac{1}{4}q\varepsilon_{k,C}^Z + q\frac{C_y^2}{\Delta_k}\mathbb{P}(C_k^{Z,M}).
\end{aligned}$$

For any $R \in [0, +\infty)$ *and any* π *such that* $C_\pi L_f^2(R_\pi \vee 1) \leq (288 C_{2.3.7}(1+T))^{-1}$*, we have, for all* $0 \leq i \leq N-1$*, that*

$$\begin{aligned}
\mathbb{E}\|y_i^R - y_i^{R,M}\|_{i,M}^2 &\leq \mathcal{E}_i^Y(R,M) + 2C_{2.3.14}^6\Big(\sum_{k=i+1}^{N-1}\Delta_k\mathcal{E}_k^Y(R,M) + \sum_{k=i}^{N-1}\Delta_k\mathcal{E}_k^Z(R,M)\Big), \\
\sum_{k=i}^{N-1}\Delta_k\mathbb{E}\|z_k^R - z_k^{R,M}\|_{k,M}^2 &\leq 2C_{2.3.14}^6\Big(\sum_{k=i+1}^{N-1}\Delta_k\mathcal{E}_k^Y(R,M) + \sum_{k=i}^{N-1}\Delta_k\mathcal{E}_k^Z(R,M)\Big).
\end{aligned}$$

Theorem 2.4.5. *Under the assumptions of Theorem 2.4.4, the bias and variance terms are bounded as follows:*

$$\begin{aligned}
T_{1,k}^{Y,M} &\leq \min_\alpha \mathbb{E}|y_k^R(X_k) - \alpha\cdot p_{0,k}(X_k)|^2, & T_{2,k}^{Y,M} &\leq C_\Psi^2\frac{K_{0,k}}{M}, \\
T_{1,k}^{Z,M} &\leq \sum_{l=1}^{q}\min_\alpha \mathbb{E}|z_{l,k}^R(X_k) - \alpha\cdot p_{l,k}(X_k)|^2, & T_{2,k}^{Z,M} &\leq \frac{C_\Psi^2}{\Delta_k}\sum_{l=1}^{q}\frac{K_{l,k}}{M}.
\end{aligned}$$

For the large deviation events, we have

$$
\begin{aligned}
\mathbb{P}(A_k^{Y,M}) \le p_{A,k}^{Y,M} :=& 2K_{0,k} \exp\big(-\frac{M\varepsilon_{k,A}^Y}{72C_\Psi^2 K_{0,k}}\big) \\
& \prod_{i=k}^{N-1} 3^{(q+1)} \Delta_i^{2(K_{0,i}+1)} \big(\frac{96K_{0,k}L_f^2 T^{1+\theta_L}(q+1)C_y^2}{\theta_L \varepsilon_{k,A}^Y \Delta_i}\big)^{2\sum_{l=0}^q (K_{l,i}+1)}, \\
\mathbb{P}(A_k^{Z,M}) \le p_{A,k}^{Z,M} :=& \sum_{l=1}^q 2K_{l,k} \exp\big(-\frac{M\Delta_k \varepsilon_{k,A}^Z}{72C_\Psi^2 R^2 K_{l,k}}\big) \\
& \prod_{i=k}^{N-1} 3^{(q+1)} \Delta_i^{2(K_{0,i}+1)} \big(\frac{96K_{l,k}L_f^2 T^{1+\theta_L}(q+1)C_y^2 R^2}{\theta_L \varepsilon_{k,A}^Z \Delta_i \Delta_k}\big)^{2\sum_{l'=0}^q (K_{l',i}+1)}, \\
\mathbb{P}(B_k^{Y,M}) \le p_{B,k}^{Y,M} :=& 4\exp\big(-\frac{M\varepsilon_{k,B}^Y}{60(C_\Psi + C_y)^2}\big) \\
& \prod_{i=k}^{N-1} 3^{(q+1)} \Delta_i^{K_{0,i}+1} \big(\frac{132 L_f T^{\frac{1+\theta_L}{2}}(q+1)(C_\Psi + C_y)C_y}{\varepsilon_{k,B}^Y \sqrt{\Delta_i}}\big)^{2\sum_{l=0}^q (K_{l,i}+1)}, \\
\mathbb{P}(B_k^{Z,M}) \le p_{B,k}^{Z,M} :=& 4\ q\exp\big(-\frac{M\varepsilon_{k,B}^Z \Delta_k}{60(C_\Psi + C_y)^2}\big) \\
& \prod_{i=k}^{N-1} 3^{(q+1)} \Delta_i^{K_{0,i}+1} \big(\frac{132 L_f T^{\frac{1+\theta_L}{2}}(q+1)(C_\Psi + C_y)C_y}{\varepsilon_{k,B}^Z \sqrt{\Delta_i}\sqrt{\Delta_k}}\big)^{2\sum_{l=0}^q (K_{l,i}+1)}, \\
\mathbb{P}(C_k^{Y,M}) \le p_{C,k}^{Y,M} :=& 12\exp\big(-\frac{M\varepsilon_{k,C}^Y}{507C_y^2}\big)\ \big(\frac{1056C_y^2}{5\varepsilon_{k,C}^Y}\big)^{2(K_{0,k}+1)}, \\
\mathbb{P}(C_k^{Z,M}) \le p_{C,k}^{Z,M} :=& 12\exp\big(-\frac{M\varepsilon_{k,C}^Y \Delta_k}{507C_y^2}\big) \sum_{l=1}^q \big(\frac{1056C_y^2}{5\varepsilon_{k,C}^Z \Delta_k}\big)^{2(K_{l,k}+1)},
\end{aligned}
$$

provided that $\varepsilon_{k,A}^Y, \varepsilon_{k,A}^Z, \varepsilon_{k,B}^Y, \varepsilon_{k,B}^Z, \varepsilon_{k,C}^Y, \varepsilon_{k,C}^Z$ *are small enough in the sense that*

$$
\begin{aligned}
& 0 < \varepsilon_{k,A}^Y \le \big(1 \wedge \min_{i=k\dots N-1} \Delta_i^{-1}\big) K_{0,k} \frac{9C_y^2 L_f^2 T^{1+\theta_L}(q+1)}{2\theta_L}, \\
& 0 < \varepsilon_{k,A}^Z \le \big(1 \wedge \min_{i=k\dots N-1} \Delta_i^{-1}\big)\big(\min_{l=1\dots q} K_{l,k}\big) \frac{R^2}{\Delta_k}\ \frac{9C_y^2 L_f^2 T^{1+\theta_L}(q+1)}{2\theta_L}, \\
& 0 < \varepsilon_{k,B}^Y \le \big(1 \wedge \min_{i=k\dots N-1} \Delta_i^{-1/2}\big) 15 L_f T^{\frac{1+\theta_L}{2}}(q+1)(C_\Psi + C_y)C_y, \\
& 0 < \varepsilon_{k,B}^Z \le \big(1 \wedge \min_{i=k\dots N-1} \Delta_i^{-1/2}\big) \frac{15 L_f T^{\frac{1+\theta_L}{2}}(q+1)(C_\Psi + C_y)C_y}{\sqrt{\Delta_k}}, \\
& 0 < \varepsilon_{k,C}^Y \le 24C_y^2, \qquad 0 < \varepsilon_{k,C}^Z \le 24C_y^2 \Delta_k^{-1/2}.
\end{aligned}
$$

2.4.3 Algorithm complexity

The error analysis of Section 2.4.2 shows us that the numerical parameters may play multiple and often contradictory roles in the convergence of the scheme: the higher the number N of steps, the smaller the discretization error but the larger effect for the propagation of errors through the DP equation; the higher the dimension of the function spaces for the empirical regression, the better the approximation accuracy (the *bias* term in regression) but the larger the statistical error (the *variance* term); the higher the number of simulations, the smaller the statistical error, but the more computational work to be done. We now demonstrate how the results of Theorems 2.4.4 and 2.4.5 can be used to optimize these parameters by means of an error vs. computational work (*complexity*) analysis. Moreover, we show how extra smoothness of the Markov functions can bring about substantial improvements in the complexity analysis.

For simplicity, we assume that the time-grid is uniform: $\Delta_i = T/N$.

Assume that the function y_i (defined in (2.1.3)) is of class $C_b^{\kappa+1+\eta}$ uniformly in i, meaning that y_i is uniformly bounded and $\kappa+1$-continuously differentiable ($\kappa \geq 0$), with bounded derivatives, and the $\kappa+1$-th derivatives are η-Hölder continuous ($\eta \in (0,1]$). Additionally, assume that z_i is $C_b^{\kappa+\eta}$. These enhanced assumptions are natural: in the continuous time case, see [CD12] for a recent account, y_i inherits the smoothness of the terminal condition and the driver, and z_i is once less differentiable that y_i. We already see in Lemma 2.3.10 and Corollary 2.3.11 that these assumptions are viable under the assumption of Lipschitz continuity for the terminal condition and the driver. The bounds on the functions y_i and z_i and their derivatives are assumed to be uniform in N. This improves the bounds in Theorems 2.4.4 and 2.4.5, because one can remove the dependence of the constant C_Ψ on N.

Let us make the squared global errors $\mathbb{E}\|y_i^R - y_i^{R,M}\|_{i,M}^2$ and $\sum_{k=i}^{N-1} \Delta_k \mathbb{E}\|z_k^R - z_k^{R,M}\|_{k,M}^2$ be of magnitude $(N^{-2\theta_{\text{conv}}})$ where $\theta_{\text{conv}} > 0$. The parameter θ_{conv} can be the convergence order of the time discretization scheme: in the case of Lipschitz continuous f and Φ in a diffusion setting, $\theta_{\text{conv}} = \frac{1}{2}$ (see [LGW06, Theorem 1]). When the forward component is simulated with a strong error of order 1, then one can achieve $\theta_{\text{conv}} = 1$ (see [GL07, Theorems 7 and 8]). It is sufficient due to Theorem 2.4.4 to tune the parameters (basis functions, number of simulations) so that each term in $\mathcal{E}_i^Y(R,M)$ and $\mathcal{E}_i^Z(R,M)$ be $O(N^{-2\theta_{\text{conv}}})$.

For the basis functions, we take local polynomials defined on disjoint hypercubes $(\mathcal{H}_n)_{n=1,\dots,K_{l,k}}$ ($l = 0,\dots,q$) with edge length δ_y (for y) and δ_z (for z_l). The union of these hypercubes is of the form $[-\bar{R},\bar{R}]^d$ for each component $y, z_1,\dots,z_q$. The degree of local polynomials is $\kappa+1$ for y, and κ for z_l. We denote by x_n the center of the n-th hypercube $\mathcal{H}_n$.

In the following, c is a positive constant that does not depend on N and may change from line to line; c is assumed to be large enough for the arguments to be consistent.

Bias terms. Because of Proposition 2.3.9, we can replace y^R and z^R by y and z (that is $R = +\infty$) in the expression of $T_{1,i}^{Y,M}$ and $T_{1,i}^{Z,M}$ by choosing $R = \sqrt{4(\theta_{\text{conv}} + \frac{1}{2})\log(N+1)}$. It adds an extra squared error $O(N^{-2\theta_{\text{conv}}})$. The projection error $\min_\alpha \mathbb{E}|y_i(X_i) - \alpha \cdot p_{0,i}(X_i)|^2$ is equal to

$$\begin{aligned}
\mathbb{E}|y_i(X_i)\mathbf{1}_{|X_i|_\infty > \bar{R}}|^2 &+ \sum_{n=1}^{K_{0,i}} \min_\alpha \mathbb{E}|y_i(X_i) - \alpha \cdot p_{0,i}(X_i)|^2 \mathbf{1}_{X_i \in \mathcal{H}_n} \\
&\leq |y_i|_\infty^2 \mathbb{P}(|X_i|_\infty > \bar{R}) + \sum_{n=1}^{K_{0,i}} c|y_i|_{\kappa+1+\eta}^2 (\delta_y^{\kappa+1+\eta})^2 \mathbb{P}(X_i \in \mathcal{H}_n) \\
&\leq |y_i|_\infty^2 \mathbb{P}(|X_i|_\infty > \bar{R}) + c|y_i|_{\kappa+1+\eta}^2 (\delta_y^{\kappa+1+\eta})^2
\end{aligned}$$

where we have used a Taylor expansion on each set $\mathcal{H}_n$ and taken the local polynomials to be equal to the first terms of the expansion. Assume additionally that X_i has exponential moments (uniformly in i), i.e. for some $\lambda > 0$, $\sup_{N\geq 1}\sup_{0\leq i\leq N}\mathbb{E}(e^{\lambda|X_i|_\infty}) < +\infty$, so that the choice $\bar{R} = 2\theta_{\text{conv}}\lambda^{-1}\log(N+1)$ is sufficient to ensure $\mathbb{P}(|X_i| > \bar{R}) = O(N^{-2\theta_{\text{conv}}})$. Hence, the choice $\delta_y = cN^{-\frac{\theta_{\text{conv}}}{\kappa+1+\eta}}$ ensures that $T_{1,i}^{Y,M} = O(N^{-2\theta_{\text{conv}}})$. With similar arguments for the z_l components, we have to choose $\delta_z = cN^{-\frac{\theta_{\text{conv}}}{\kappa+\eta}}$. Thus the sizes of the vector spaces are $K_{0,i} = cN^{d\frac{\theta_{\text{conv}}}{\kappa+1+\eta}}\log^d(N+1)$ and $K_{l,i} = cN^{d\frac{\theta_{\text{conv}}}{\kappa+\eta}}\log^d(N+1)$. Observe that, for the above analysis, we could alternatively assume that the continuous-time BSDE (Y,Z) associated to $(y_i(X_i), z_i(X_i))_i$ has a Markovian representation $Y_t = u(t, X_t)$ and $Z_t = v(t, X_t)$, where u and v are respectively of class $C_b^{\kappa+1+\eta}$ and $C_b^{\kappa+\eta}$ in space. Then, observing that

$$\min_\alpha \mathbb{E}|y_i(X_i) - \alpha\cdot p_{0,i}(X_i)|^2 \leq \min_\alpha 2\mathbb{E}|u(t_i, X_{t_i}) - \alpha\cdot p_{0,i}(X_i)|^2 + O(N^{-2\theta_{\text{conv}}})$$

and similarly for Z, we are reduced to the previous analysis and we can achieve the same conclusion regarding the tuning of $K_{l,i}$, $l = 0, \ldots, q$. This observation may be very useful if estimates on the true BSDE (Y,Z) or on its semi-linear value function (u,v) are known explicitly (like in [CD12]).

Variance terms. Making $T_{2,i}^{Z,M}$ of order $N^{-2\theta_{\text{conv}}}$ implies $M = cN^{1+2\theta_{\text{conv}}}K_{l,i} = cN^{1+2\theta_{\text{conv}}+d\frac{\theta_{\text{conv}}}{\kappa+\eta}}\log^d(N+1)$; this dominates the requirements on M for $T_{2,i}^{Y,M}$.

Large deviation events. We set $\varepsilon_{k,A}^Y, \varepsilon_{k,B}^Y, \varepsilon_{k,C}^Y, \varepsilon_{k,A}^Z, \varepsilon_{k,B}^Z, \varepsilon_{k,C}^Z$ equal to $N^{-2\theta_{\text{conv}}}$. In order to make the probability upper bound exponentially small a quick look at Theorem 2.4.5 shows that the strongest constraint comes from $\mathbb{P}(A_i^{Z,M})$ which imposes $c(NK_{l,i}\log(N+1) + \log(N+1)) = \frac{M}{N^{1+2\theta_{\text{conv}}}K_{l,i}\log(N+1)}$. This condition on M is much stronger than the previous one, ensuring that the variance terms have the right magnitude; the requirement of having more simulations seems to be the price to pay for having a single set of paths and interdependent regression problems. To fulfill this condition, take

$$M = cK_{l,i}^2 N^{2+2\theta_{\text{conv}}}\log^2(N+1) = cN^{2+2\theta_{\text{conv}}+2d\frac{\theta_{\text{conv}}}{\kappa+\eta}}\log^{2+2d}(N+1).$$

Complexity analysis for the LSMDP scheme. Due to the properties of hypercubes (disjoint intervals), the final computational cost $\mathcal{C}$ (counting the elementary operations) is of order MN, that is

$$\mathcal{C} = cN^{3+2\theta_{\text{conv}}+2d\frac{\theta_{\text{conv}}}{\kappa+\eta}}\log^{2+2d}(N+1).$$

Equivalently, the global error, as a function of complexity and ignoring log factors, is

$$N^{-\theta_{\text{conv}}} \leq c\,\mathcal{C}^{\frac{-\theta_{\text{conv}}}{3+2\theta_{\text{conv}}+2d\frac{\theta_{\text{conv}}}{\kappa+\eta}}} = c\,\mathcal{C}^{\frac{-1}{2(1+\frac{3}{2\theta_{\text{conv}}}+\frac{d}{\kappa+\eta})}}. \tag{2.4.22}$$

This analysis shows that the smaller the parameter $\frac{3}{2\theta_{\text{conv}}} + \frac{d}{\kappa+\eta}$, the quicker the convergence. There are several numerically significant implications of this:

- The higher the smoothness of the solution, the better the convergence. This may motivate first solving the BSDE without driver and then solve the BSDE difference (which in general gives a smoother problem, see [GM10]) - see our discussion about proxys in the Section 2.2.
- The higher the dimension, the worse the convergence. This is the usual curse of dimensionality.

- The better the discretization error (θ_{conv} large), the better the convergence. This motivates the development of high-order discretization schemes for BSDEs.

If we apply the same analysis to [LGW06, Theorem 2] for $\kappa + \eta \geq 1$ and $\theta_{\text{conv}} = 1/2$, we obtain that the error is of order $\mathcal{C}^{-\frac{1}{2(4+\frac{2d}{\kappa+1+\eta})}}$ for the ODP. In contrast, the LSMDP has error of order $\mathcal{C}^{-\frac{1}{2(4+\frac{d}{\kappa+\eta})}}$. This implies that, for sufficiently large N, the MDP performs better than the ODP for $\kappa + \eta > 1$, at least in the theoretical framework given in this section. For $\kappa + \eta = 1$, the performance is the same. In the future, we will undertake numerical comparisons between ODP and MDP to test this theoretical analysis.

2.4.4 Proof of Theorem 2.4.4

From $[y_k^R]_y = y_k^R$ and the Lipschitz property of $[\cdot]_y$, it follows that

$$\mathbb{E}\|y_k^R - y_k^{R,M}\|_{k,M}^2 \leq \mathbb{E}\|y_k^R - \alpha_{0,k}^M \cdot p_{0,k}\|_{k,M}^2.$$

For any m,

$$\mathbb{E}_N^M\Big[\Phi(\tilde{X}_N^{k,m}) + \sum_{i=k}^{N-1} f_i(\tilde{X}_i^{k,m}, y_{i+1}^R(\tilde{X}_{i+1}^{k,m}), z_i^R(\tilde{X}_i^{k,m}))\Delta_i\Big] = y_k^R(\tilde{X}_k^{k,m}) = y_k^R(X_k^m)$$

is clear from the definition of y^R in (2.4.1). Hence, owing to (2.4.10) and Proposition 2.4.1(iii) (applied with $\mathcal{Q} = \mathcal{F}_N^M$), $\mathbb{E}_N^M[\tilde{\beta}_{0,k}^M]$ is the SVD-minimizer of $\|y_k^R - \alpha \cdot p_{0,k}\|_{k,M}^2$. We now apply Pythagoras' theorem and take expectations to obtain

$$\mathbb{E}\|y_k^R - y_k^{R,M}\|_{k,M}^2 \leq T_{1,k}^{Y,M} + \mathbb{E}\|(\alpha_{0,k}^M - \mathbb{E}_N^M[\tilde{\beta}_{0,k}^M]) \cdot p_{0,k}\|_{k,M}^2.$$

To decompose the last term above, we introduce the coefficients $\tilde{\alpha}_{0,k}^M$ and $\mathbb{E}_N^M(\tilde{\alpha}_{0,k}^M)$:

$$\begin{aligned}
&\mathbb{E}\|(\alpha_{0,k}^M - \mathbb{E}_N^M[\tilde{\beta}_{0,k}^M]) \cdot p_{0,k}\|_{k,M}^2 \\
&\quad \leq 3\mathbb{E}\|(\alpha_{0,k}^M - \tilde{\alpha}_{0,k}^M) \cdot p_{0,k}\|_{k,M}^2 + 3T_{2,k}^{Y,M} + 3\mathbb{E}\|(\mathbb{E}_N^M[\tilde{\beta}_{0,k}^M - \tilde{\alpha}_{0,k}^M]) \cdot p_{0,k}\|_{k,M}^2.
\end{aligned}$$

The first term on the r.h.s. is estimated using the event $A_k^{Y,M}$ in (2.4.15). To do this, we first need to obtain an almost sure bound on the integrand. Indeed, from Proposition 2.4.1(i), $\alpha_{0,k}^M - \tilde{\alpha}_{0,k}^M$ is the SVD-optimal coefficient of the least-squares problem w.r.t. the $\|\cdot\|_{k,M}$-norm associated to the differences of the responses of $\alpha_{0,k}^M$ and $\tilde{\alpha}_{0,k}^M$. Proposition 2.4.1(ii) then applies to give $\|(\alpha_{0,k}^M - \tilde{\alpha}_{0,k}^M) \cdot p_{0,k}\|_{k,M}^2 \leq 4C_\Psi^2$. Using this upper bound, we now have

$$\mathbb{E}\|(\alpha_{0,k}^M - \tilde{\alpha}_{0,k}^M) \cdot p_{0,k}\|_{k,M}^2 \leq \varepsilon_{k,A}^Y + 4C_\Psi^2 \mathbb{P}(A_k^{Y,M}). \tag{2.4.23}$$

To handle $\mathbb{E}\|(\mathbb{E}_N^M[\tilde{\beta}_{0,k}^M - \tilde{\alpha}_{0,k}^M]) \cdot p_{0,k}\|_{k,M}^2$, observe that $\mathbb{E}_N^M[\tilde{\beta}_{0,k}^M - \tilde{\alpha}_{0,k}^M]$ is the SVD-optimal coefficient of the least-squares problem related to the response

$$
\begin{aligned}
&\mathbb{E}_N^M\Big[\Phi(\tilde{X}_N^{k,m}) + \sum_{i=k}^{N-1} f_i(\tilde{X}_i^{k,m}, y_{i+1}^R(\tilde{X}_{i+1}^{k,m}), z_i^R(\tilde{X}_i^{k,m}))\Delta_i\Big] \\
&- \mathbb{E}_N^M\Big[\Phi(\tilde{X}_N^{k,m}) + \sum_{i=k}^{N-1} f_i(\tilde{X}_i^{k,m}, y_{i+1}^{R,M}(\tilde{X}_{i+1}^{k,m}), z_i^{R,M}(\tilde{X}_i^{k,m}))\Delta_i\Big] \\
&= y_k^R(X_k^m) - \bar{y}_k^{R,M}(X_k^m) \quad \text{(by Lemma 2.4.3).}
\end{aligned}
$$

By the contraction property (item (ii) of Proposition 2.4.1),

$$
\begin{aligned}
\mathbb{E}\|(\mathbb{E}_N^M[\tilde{\beta}_{0,k}^M - \tilde{\alpha}_{0,k}^M]) \cdot p_{0,k}\|_{k,M}^2 &\le \mathbb{E}\|y_k^R - \bar{y}_k^{R,M}\|_{k,M}^2 \\
&\le \varepsilon_{k,B}^Y + 2\mathbb{E}(y_k^R(X_k) - \bar{y}_k^{R,M}(X_k))^2 + 2(C_y^2 + C_\Psi^2)\mathbb{P}(B_k^{Y,M}).
\end{aligned}
$$

Bringing together the thus far obtained results yields

$$
\begin{aligned}
\mathbb{E}\|y_k^R - y_k^{R,M}\|_{k,M}^2 \le\ & T_{1,k}^{Y,M} + 3\varepsilon_{k,A}^Y + 12C_\Psi^2\mathbb{P}(A_k^{Y,M}) \\
&+ 3T_{2,k}^{Y,M} + 3\varepsilon_{k,B}^Y + 6(C_y^2 + C_\Psi^2)\mathbb{P}(B_k^{Y,M}) \\
&+ 6\mathbb{E}[(y_k^R(X_k) - \bar{y}_k^{R,M}(X_k))^2].
\end{aligned} \tag{2.4.24}
$$

We can perform analogous calculations for the Z component (replacing C_Ψ by $\frac{C_\Psi R}{\sqrt{\Delta_k}}$ for the $A_k^{Z,M}$-event, C_Ψ by $\frac{C_\Psi}{\sqrt{\Delta_k}}$ for the $B_k^{Z,M}$-event, and C_y by $\frac{C_y}{\sqrt{\Delta_k}}$), obtaining

$$
\begin{aligned}
\mathbb{E}\|z_k^R - z_k^{R,M}\|_{k,M}^2 \le\ & T_{1,k}^{Z,M} + 3q\varepsilon_{k,A}^Z + 12q\frac{C_\Psi^2 R^2}{\Delta_k}\mathbb{P}(A_k^{Z,M}) \\
&+ 3T_{2,k}^{Z,M} + 3q\varepsilon_{k,B}^Z + 6q\frac{(C_y^2 + C_\Psi^2)}{\Delta_k}\mathbb{P}(B_k^{Z,M}) \\
&+ 6\mathbb{E}[|z_k^R(X_k) - \bar{z}_k^{R,M}(X_k)|^2].
\end{aligned} \tag{2.4.25}
$$

Observe that here, we rely on truncated Brownian increments in order to *a.s.* upper bound $\|(\alpha_{l,k}^M - \tilde{\alpha}_{l,k}^M) \cdot p_{l,k}\|_{k,M}^2$ by $4\frac{C_\Psi^2 R^2}{\Delta_k}$. We can use the same reasoning as in Lemma 2.4.3 to show that

$$
\begin{aligned}
\bar{y}_k^{R,M}(X_k) =& \mathbb{E}[\Phi(X_N) + \sum_{i=k}^{N-1} f_i(X_i, y_{i+1}^{R,M}(X_{i+1}), z_i^{R,M}(X_i))\Delta_i \mid \mathcal{F}_N^M \vee \mathcal{F}_{t_k}], \\
\Delta_k \bar{z}_{l,k}^{R,M}(X_k) =& \mathbb{E}\big[[\Delta W_{l,k}]_w\big(\Phi(X_N) \\
&+ \sum_{i=k+1}^{N-1} f_i(X_i, y_{i+1}^{R,M}(X_{i+1}), z_i^{R,M}(X_i))\Delta_i\big) \mid \mathcal{F}_N^M \vee \mathcal{F}_{t_k}\big],
\end{aligned}
$$

which implies that $(\bar{y}_k^{R,M}(X_k), \bar{z}_k^{R,M}(X_k))_{k\ge 0}$ solves a discrete BSDE with data $(\Phi(X_N), \bar{f}_i(y,z) = f_i(X_i, y_{i+1}^{R,M}(X_{i+1}), z_i^{R,M}(X_i)))$. We now apply the stability result from Proposition 2.3.3 (w.r.t. filtration $(\mathcal{F}_N^M \vee \mathcal{F}_{t_k})_{0\le k\le N}$), taking the first discrete BSDE to be $(y_k^R(X_k), z_k^R(X_k))_{k\ge 0}$ and the second to be $(\bar{y}_k^{R,M}(X_k), \bar{z}_k^{R,M}(X_k))_{k\ge 0}$ ($L_{\bar{f}_k} = 0$). Combined with the local Lipschitz continuity of f_k and a choice of $\gamma \in (0, +\infty)^N$ such that

$$144(R_\pi \vee 1)C_{2.3.7}(1+T)(\frac{1}{\gamma_k} + \Delta_k)\frac{L_f^2}{(T-t_k)^{1-\theta_L}} \leq 1, \qquad (0 \leq k < N), \tag{2.4.26}$$

we obtain the bound

$$\begin{aligned}
&\sum_{k=i+1}^{N-1} \Delta_k \mathbb{E}[|y_k^R(X_k) - \bar{y}_k^{R,M}(X_k)|^2]\Gamma_k + \sum_{k=i}^{N-1} \Delta_k \mathbb{E}[|z_k^R(X_k) - \bar{z}_k^{R,M}(X_k)|^2]\Gamma_k \\
&\qquad \leq 6C_{2.3.7}(1+T)\sum_{k=i}^{N-1}(\frac{1}{\gamma_k} + \Delta_k)\Delta_k \frac{L_f^2}{(T-t_k)^{1-\theta_L}} \\
&\qquad\qquad \times \mathbb{E}\big[|y_{k+1}^R(X_{k+1}) - y_{k+1}^{R,M}(X_{k+1})|^2 + |z_k^R(X_k) - z_k^{R,M}(X_k)|^2\big]\Gamma_k \\
&\qquad \leq \frac{1}{12}\mathcal{E}_{i,N}(\gamma, R, M) \qquad\qquad (2.4.27)\\
&\qquad + \frac{1}{24}\sum_{k=i}^{N-1}\Delta_k\big(1_{k+1<N}[\varepsilon_{k+1,C}^Y + 4C_y^2\mathbb{P}(C_{k+1}^{Y,M})] + q\varepsilon_{k,C}^Z + 4q\frac{C_y^2}{\Delta_k}\mathbb{P}(C_k^{Z,M})\big)\Gamma_k
\end{aligned}$$

where

$$\mathcal{E}_{i,N}(\gamma, R, M) := \sum_{k=i+1}^{N-1}\Delta_k \mathbb{E}\|y_k^R - y_k^{R,M}\|_{k,M}^2\Gamma_k + \sum_{k=i}^{N-1}\Delta_k\mathbb{E}\|z_k^R - z_k^{R,M}\|_{k,M}^2\Gamma_k.$$

Using (2.4.27) together with (2.4.24) and (2.4.25), it readily follows that

$$\begin{aligned}
\mathcal{E}_{i,N}(\gamma, R, M) \leq &\sum_{k=i+1}^{N-1}\Delta_k\{T_{1,k}^{Y,M} + 3\varepsilon_{k,A}^Y + 12C_\Psi^2\mathbb{P}(A_k^{Y,M}) + 3T_{2,k}^{Y,M} + 3\varepsilon_{k,B}^Y \\
&\quad + 6(C_y^2 + C_\Psi^2)\mathbb{P}(B_k^{Y,M})\}\Gamma_k \\
&+ \sum_{k=i}^{N-1}\Delta_k\{T_{1,k}^{Z,M} + 3q\varepsilon_{k,A}^Z + 12q\frac{C_\Psi^2 R^2}{\Delta_k}\mathbb{P}(A_k^{Z,M}) + 3T_{2,k}^{Z,M} + 3q\varepsilon_{k,B}^Z \\
&+ 6q\frac{(C_y^2 + C_\Psi^2)}{\Delta_k}\mathbb{P}(B_k^{Z,M})\}\Gamma_k + \frac{1}{2}\mathcal{E}_{i,N}(\gamma, R, M) \\
&+ \frac{1}{4}\sum_{k=i}^{N-1}\Delta_k\{1_{k+1<N}[\varepsilon_{k+1,C}^Y + 4C_y^2\mathbb{P}(C_{k+1}^{Y,M})] + q\varepsilon_{k,C}^Z + 4q\frac{C_y^2}{\Delta_k}\mathbb{P}(C_k^{Z,M})\}\Gamma_k.
\end{aligned}$$

The choice $\gamma_k = 288(R_\pi \vee 1)C_{2.3.7}(1+T)\frac{L_f^2}{(T-t_k)^{1-\theta_L}}$ leads to $1 \leq \Gamma_i \leq C_{2.3.14}^6$ and then for N large enough (such that $C_\pi L_f^2(R_\pi \vee 1) \leq \frac{1}{288C_{2.3.7}(1+T)}$), the condition (2.4.26) is satisfied and we obtain the announced estimate on the Z component.
Now, applying the same arguments used in (2.4.27) directly to $\mathbb{E}[|y_i^R(X_i) - y_i^{R,M}(X_i)|^2]$ in (2.4.24),

we obtain

$$\mathbb{E}\|y_i^R - y_i^{R,M}\|_{i,M}^2 \Gamma_i \le \mathcal{E}_i^Y(R,M)\Gamma_i + \frac{1}{2}\mathcal{E}_{i,N}(\gamma, R, M) + \sum_{k=i+1}^{N-1} \Delta_k \mathcal{E}_k^Y(R,M)\Gamma_k + \sum_{k=i}^{N-1} \Delta_k \mathcal{E}_k^Z(R,M)\Gamma_k$$

and we easily complete the proof for the Y component. □

2.4.5 Proof of Theorem 2.4.5

In each of the following subsections, we prove the bounds for the local error terms given in Theorem 2.4.5. **Bias/variance terms**The bounds on squared bias terms $T_{1,k}^{Y,M}$ and $T_{1,k}^{Z,M}$ are straightforward. For the variance terms $T_{2,k}^{Y,M}$ and $T_{2,k}^{Z,M}$, we use the same arguments as [GKKW02, pp.186–187] or [LGW06, Proposition 4]. □

Large deviation events In the proofs below we use *covering* techniques to allow us to apply exponential inequalities to the large deviation events. We recall few definitions here for the benefit of the reader; for a fuller account, see [GKKW02, Chapter 9].
If $\mathcal{G}$ is a class of functions from $\mathbb{R}^d$ to $\mathbb{R}$ and $x^{1:M} = \{x^m\}_{m=1\dots M}$ are M points of $\mathbb{R}^d$, an ε-*cover* ($\varepsilon > 0$) of $\mathcal{G}$ w.r.t. the $\mathbf{L}_p(p \ge 1)$-empirical norm $\|g\|_M = \big(\frac{1}{M}\sum_{m=1}^M |g(x^m)|^p\big)^{\frac{1}{p}}$ is a finite collection of functions $g_1, \dots, g_n \in \mathcal{G}$ such that for any $g \in \mathcal{G}$, we can find a $j \in \{1, \cdots, n\}$ such that $\|g - g_j\|_M \le \varepsilon$. The smallest integer n for which an ε-cover exists is called the *ε-covering number* and denoted by $\mathcal{N}_p(\varepsilon, \mathcal{G}, x^{1:M})$; we usually consider an ε-cover with minimal number of elements. In the following the points $x^{1:M}$ are possibly random. More generally, we may consider ε-cover w.r.t. the $\mathbf{L}_p$-norm of a probability measure ν (instead of the empirical measure associated to $x^{1:M}$): the related covering number is then denoted $\mathcal{N}_p(\varepsilon, \mathcal{G}, \nu)$.

Bounds on $\mathbb{P}(A_k^{Y,M})$ and $\mathbb{P}(A_k^{Z,M})$ We only prove the bound related to $A_k^{Y,M}$; the proof for $A_k^{Z,M}$ is analogous. We use a similar method to [LGW06] but with some important differences. Define the following sets of functions

$$\begin{cases} [\mathcal{S}_N^Y]_y := \{\Phi\}, \qquad [\mathcal{S}_k^Y]_y := \big\{[\alpha \cdot p_{0,k}]_y : \alpha \in \mathbb{R}^{K_{0,k}}\big\}, \\ [\mathcal{S}_k^{Z,l}]_z := \big\{[\alpha \cdot p_{l,k}]_z : \alpha \in \mathbb{R}^{K_{l,k}}\big\} \qquad (1 \le l \le q), \end{cases} \tag{2.4.28}$$

$$\mathcal{G}_k := \Big\{(x_k, \dots, x_N) \in \bigotimes_{i=k}^N \mathbb{R}^d \mapsto \Phi(x_N) + \sum_{i=k}^{N-1} f_i(x_i, \psi_{i+1}(x_{i+1}), \eta_i(x_i))\Delta_i : \psi_i \in [\mathcal{S}_i^Y]_y, \quad \eta_{l,i} \in [\mathcal{S}_i^{Z,l}]_z\Big\}$$

for $k = N-1, \dots, 0$. These definitions originate from the observation that $y_k^{R,M} \in [\mathcal{S}_k^Y]_y$, $z_{l,k}^{R,M} \in [\mathcal{S}_k^{Z,l}]_z$ and $\Psi_k^{R,M} \in \mathcal{G}_k$. By the same arguments as in Lemma 2.4.2, every element in $\mathcal{G}_k$ is bounded by C_Ψ. Define

$$G_k^m := \Psi_k^{R,M}(X_k^m, \dots, X_N^m), \qquad \tilde{G}_k^m := \Psi_k^{R,M}(\tilde{X}_k^{k,m}, \dots, \tilde{X}_N^{k,m}).$$

Similarly, for any $G \in \mathcal{G}_k$, we write G^m or $\tilde{G}^m$ for the function evaluated along the m-th sample path or its ghost path. Firstly, observe that

$$\mathbb{P}\big(\|(\alpha_{0,k}^M - \tilde{\alpha}_{0,k}^M) \cdot p_{0,k}\|_{k,M}^2 > \varepsilon_{k,A}^Y\big) = \mathbb{E}\big(\mathbb{P}_k^M\big(\|(\alpha_{0,k}^M - \tilde{\alpha}_{0,k}^M) \cdot p_{0,k}\|_{k,M}^2 > \varepsilon_{k,A}^Y\big)\big).$$

Conditionally on $\mathcal{F}_k^M$, we can assume that the basis functions are orthonormalized w.r.t. $\langle \cdot, \cdot \rangle_{k,M}$ and that the coefficients $\alpha_{0,k}^M$ and $\tilde{\alpha}_{0,k}^M$ are computed for the orthonormalized functions. Nevertheless, the dimension of the vector space w.r.t. $\langle \cdot, \cdot \rangle_{k,M}$ may be smaller than $K_{0,k}$; we denote this empirical dimension of the vector space by $K_{0,k}^M$. Then

$$\|(\alpha_{0,k}^M - \tilde{\alpha}_{0,k}^M) \cdot p_{0,k}\|_{k,M}^2 = |\alpha_{0,k}^M - \tilde{\alpha}_{0,k}^M|^2_{\mathbb{R}^{K_{0,k}^M}}.$$

Furthermore, the coefficients $\alpha_{0,k}^M$ and $\tilde{\alpha}_{0,k}^M$ have simple expressions and we obtain

$$\begin{aligned}
\mathbb{P}_k^M(|\alpha_{0,k}^M - \tilde{\alpha}_{0,k}^M|^2_{\mathbb{R}^{K_{0,k}^M}} > \varepsilon_{k,A}^Y) &= \mathbb{P}_k^M\Big(\sum_{i=1}^{K_{0,k}^M} \Big|\frac{1}{M}\sum_{m=1}^M p_{0,k}^{i,m}(G_k^m - \tilde{G}_k^m)\Big|^2 > \varepsilon_{k,A}^Y\Big) \\
&\le \sum_{i=1}^{K_{0,k}^M} \mathbb{P}_k^M\Big(\Big|\frac{1}{M}\sum_{m=1}^M p_{0,k}^{i,m}(G_k^m - \tilde{G}_k^m)\Big|^2 > \frac{\varepsilon_{k,A}^Y}{K_{0,k}^M}\Big) \\
&\le \sum_{i=1}^{K_{0,k}^M} \mathbb{P}_k^M\Big(\Big|\frac{1}{M}\sum_{m=1}^M p_{0,k}^{i,m}(G_k^m - \tilde{G}_k^m)\Big|^2 > \frac{\varepsilon_{k,A}^Y}{K_{0,k}}\Big) \\
&\le \sum_{i=1}^{K_{0,k}^M} \mathbb{P}_k^M\Big(\exists G \in \mathcal{G}_k : \Big|\frac{1}{M}\sum_{m=1}^M p_{0,k}^{i,m}(G^m - \tilde{G}^m)\Big|^2 > \frac{\varepsilon_{k,A}^Y}{K_{0,k}}\Big) \\
&= \sum_{i=1}^{K_{0,k}^M} \mathbb{P}_k^M\Big(\exists G \in \mathcal{G}_k : \Big|\frac{1}{M}\sum_{m=1}^M p_{0,k}^{i,m}U^m(G^m - \tilde{G}^m)\Big|^2 > \frac{\varepsilon_{k,A}^Y}{K_{0,k}}\Big)
\end{aligned}$$

where U^m are independent random variables uniformly distributed on $\{-1, +1\}$ that are also independent of everything else. The last equality follows by an invariance of the $\mathbb{P}_k^M$-distribution of $G^m - \tilde{G}^m$ under multiplication by -1.

We now introduce a particular cover for the set $\mathcal{G}_k$. For a function $g : \mathbb{R}^d \mapsto \mathbb{R}$, we define the squared $\mathbf{L}_2$ empirical norm

$$\|g\|_{i,M,\tilde{M}}^2 := \frac{1}{2M}\sum_{m=1}^M \big(|g(X_i^m)|^2 + |g(\tilde{X}_i^{k,m})|^2\big).$$

Then for $\varepsilon > 0$ (the value of which is chosen later), denote by $[\mathcal{S}_i^Y(\varepsilon)]_y$ and $[\mathcal{S}_i^{Z,l}(\varepsilon)]_z$ ε-covers for $[\mathcal{S}_i^Y]_y$ and $[\mathcal{S}_i^{Z,l}]_z$, respectively, w.r.t. the norm $\|\cdot\|_{i,M,\tilde{M}}$. Take

$$\begin{aligned}
\mathcal{G}_k(\varepsilon) := \Big\{(x_k, \ldots, x_N) \mapsto \Phi(x_N) + \sum_{i=k}^{N-1} f_i(x_i, g_{i+1}^Y(x_{i+1}), g_i^Z(x_i))\Delta_i \\
: \ g_i^Y \in [\mathcal{S}_i^Y(\varepsilon)]_y,\ g_{l,i}^Z \in [\mathcal{S}_i^{Z,l}(\varepsilon)]_z\Big\}.
\end{aligned}$$

Since elements of $[\mathcal{S}_i^Y(\varepsilon)]_y$ and $[\mathcal{S}_i^{Z,l}(\varepsilon)]_z$ are bounded by C_y and $C_{z,i}$, elements of $\mathcal{G}_k(\varepsilon)$ are bounded by C_Ψ.
For every $G \in \mathcal{G}_k$, there exists a $G_\varepsilon \in \mathcal{G}_k(\varepsilon)$ such that

$$\begin{aligned}
&\frac{1}{2M}\sum_{m=1}^{M}\big\{|G^m - G_\varepsilon^m|^2 + |\tilde{G}^m - \tilde{G}_\varepsilon^m|^2\big\} \\
&= \frac{1}{2M}\sum_{m=1}^{M}\big\{\big|\sum_{i=k}^{N-1}\Delta_i\big(f_i(X_i^m, \psi_{i+1}(X_{i+1}^m), \eta_i(X_i^m)) - f_i(X_i^m, g_{i+1}^Y(X_{i+1}^m), g_i^Z(X_i^m))\big)\big|^2 \\
&\quad + \big|\sum_{i=k}^{N-1}\Delta_i\big(f_i(\tilde{X}_i^{k,m}, \psi_{i+1}(\tilde{X}_{i+1}^{k,m}), \eta_i(\tilde{X}_i^{k,m})) - f_i(\tilde{X}_i^{k,m}, g_{i+1}^Y(\tilde{X}_{i+1}^{k,m}), g_i^Z(\tilde{X}_i^{k,m}))\big)\big|^2\big\} \\
&\le \frac{1}{2M}\sum_{m=1}^{M} 2T\sum_{i=k}^{N-1}\frac{\Delta_i L_f^2}{(T-t_i)^{1-\theta_L}}\big\{|\psi_{i+1}(X_{i+1}^m) - g_{i+1}^Y(X_{i+1}^m)|^2 + |\eta_i(X_i^m) - g_i^Z(X_i^m)|^2 \\
&\qquad + |\psi_{i+1}(\tilde{X}_{i+1}^{k,m}) - g_{i+1}^Y(\tilde{X}_{i+1}^{k,m})|^2 + |\eta_i(\tilde{X}_i^{k,m}) - g_i^Z(\tilde{X}_i^{k,m})|^2\big\} \\
&\le 2\frac{T^{1+\theta_L}}{\theta_L}L_f^2(q+1)\varepsilon^2
\end{aligned}$$

where we have used $(\mathbf{A_F})$. Taking $\varepsilon^2 = \frac{\theta_L \varepsilon_{k,A}^Y}{18K_{0,k}L_f^2 T^{1+\theta_L}(q+1)}$ and $G_\varepsilon \in \mathcal{G}_k(\varepsilon)$ as above, we easily obtain

$$\begin{aligned}
\big|\frac{1}{M}\sum_{m=1}^{M} p_{0,k}^{i,m}U^m(G^m - \tilde{G}^m)\big|^2 &\le 3\big|\frac{1}{M}\sum_{m=1}^{M} p_{0,k}^{i,m}U^m(G_\varepsilon^m - \tilde{G}_\varepsilon^m)\big|^2 \\
&\quad + 3\big(\frac{1}{M}\sum_{m=1}^{M}|p_{0,k}^{i,m}|^2\big)\big(\frac{1}{M}\sum_{m=1}^{M}\big\{|G^m - G_\varepsilon^m|^2 + |\tilde{G}_\varepsilon^m - \tilde{G}^m|^2\big\}\big) \\
&\le 3\big|\frac{1}{M}\sum_{m=1}^{M} p_{0,k}^{i,m}U^m(G_\varepsilon^m - \tilde{G}_\varepsilon^m)\big|^2 + \frac{2}{3}\frac{\varepsilon_{k,A}^Y}{K_{0,k}}
\end{aligned}$$

where we take advantage of the orthonormality property of $p_{0,k}^i$. Then it follows

$$\begin{aligned}
&\mathbb{P}_k^M\big(\exists G \in \mathcal{G}_k : \big|\frac{1}{M}\sum_{m=1}^{M} p_{0,k}^{i,m}U^m(G^m - \tilde{G}^m)\big|^2 > \frac{\varepsilon_{k,A}^Y}{K_{0,k}}\big) \\
&\le \mathbb{P}_k^M\big(\exists G_\varepsilon \in \mathcal{G}_k(\varepsilon) : \big|\frac{1}{M}\sum_{m=1}^{M} p_{0,k}^{i,m}U^m(G_\varepsilon^m - \tilde{G}_\varepsilon^m)\big| > \frac{1}{3}\big(\frac{\varepsilon_{k,A}^Y}{K_{0,k}}\big)^{\frac{1}{2}}\big) \\
&\le \mathbb{E}_k^M|\mathcal{G}_k(\varepsilon)|\max_{G_\varepsilon \in \mathcal{G}_k(\varepsilon)}\tilde{\mathbb{P}}^M\big(\big|\frac{1}{M}\sum_{m=1}^{M} p_{0,k}^{i,m}U^m(G_\varepsilon^m - \tilde{G}_\varepsilon^m)\big| > \frac{1}{3}\big(\frac{\varepsilon_{k,A}^Y}{K_{0,k}}\big)^{\frac{1}{2}}\big) \\
&\le 2\mathbb{E}_k^M|\mathcal{G}_k(\varepsilon)|\exp\big(-\frac{2M}{\frac{1}{M}\sum_{m=1}^{M}|4C_\Psi p_{0,k}^{i,m}|^2} \times \frac{\varepsilon_{k,A}^Y}{9K_{0,k}}\big) = 2\mathbb{E}_k^M|\mathcal{G}_k(\varepsilon)|\exp\big(-\frac{M\varepsilon_{k,A}^Y}{72C_\Psi^2 K_{0,k}}\big)
\end{aligned}$$

where —$\mathcal{G}_k(\varepsilon)|$ denotes the number of elements of $\mathcal{G}_k(\varepsilon)$, $\tilde{\mathbb{P}}^M$ is the conditional probability with respect to $\mathcal{F}_N^M \vee \sigma(\tilde{X}_i^{k,m} : i \ge k, m = 1, \dots, M)$, and the final inequality follows from Hoeffding's inequality [GKKW02, Lemma A.3].

It remains to bound $|\mathcal{G}_k(\varepsilon)|$, which is equal to the product of the ε-covering numbers of $[\mathcal{S}_i^Y(\varepsilon)]_y$ $(i = k+1, \ldots, N-1)$ and $[\mathcal{S}_i^{Z,l}(\varepsilon)]_z$ $(l = 1, \ldots, q, i = k, \ldots, N-1)$ w.r.t. $\|\cdot\|_{i,M,\tilde{M}}$. From [GKKW02, inequality (9.22) pp.153], we have

$$\mathcal{N}_2(\varepsilon, [\mathcal{S}_i^Y(\varepsilon)]_y, [X_i^{1:M}, \tilde{X}_i^{k,1:M}]) \leq 3\big(2e(\frac{2C_y}{\varepsilon})^2 \log(3e(\frac{2C_y}{\varepsilon})^2)\big)^{K_{0,i}+1} \leq 3\big(\frac{16C_y^2}{3\varepsilon^2}\big)^{2(K_{0,i}+1)}$$

for any $\varepsilon \leq C_y/2$; the second inequality follows from the concavity of $\log(\cdot)$:

$$\forall x \geq 16, \quad 2e\,x\log(3e\,x) \leq 2e\,x[\log(48e) + \frac{3e\,x - 48e}{48e}] \leq e\frac{1+\log(48)}{8}x^2 \leq (\frac{4}{3}x)^2.$$

A similar inequality holds for $[\mathcal{S}_i^{Z,l}(\varepsilon)]_z$ by replacing C_y by $C_y/\sqrt{\Delta_i}$ and $K_{0,i}$ and $K_{l,i}$. Finally, we obtain

$$\begin{aligned}
|\mathcal{G}_k(\varepsilon)| &\leq \prod_{i=k}^{N-1} 3^{(q+1)}\big(\frac{16C_y^2}{3\varepsilon^2}\big)^{2(K_{0,i}+1)}\big(\frac{16C_y^2}{3\varepsilon^2\Delta_i}\big)^{2\sum_{l=1}^q(K_{l,i}+1)} \\
&= \prod_{i=k}^{N-1} 3^{(q+1)}\Delta_i^{2(K_{0,i}+1)}\big(\frac{96K_{0,k}L_f^2T^{1+\theta_L}(q+1)C_y^2}{\theta_L\varepsilon_{k,A}^Y\Delta_i}\big)^{2\sum_{l=0}^q(K_{l,i}+1)}
\end{aligned}$$

provided that $\varepsilon^2 = \frac{\theta_L\varepsilon_{k,A}^Y}{18K_{0,k}L_f^2T^{1+\theta_L}(q+1)} \leq \frac{C_y^2}{4} \wedge \frac{C_y^2}{4\Delta_k} \wedge \cdots \wedge \frac{C_y^2}{4\Delta_{N-1}}$. Gather different inequalities and bound $K_{0,k}^M$ by $K_{0,k}$ to derive the announced result. □

Bounds on $\mathbb{P}(B_k^{Y,M})$ and $\mathbb{P}(B_k^{Z,M})$ Again, we only detail the proof for $B_k^{Y,M}$. We define the set of functions

$$\begin{aligned}
\mathcal{G}'_k := \big\{x \in \mathbb{R}^d \mapsto \int \big(\Phi(x_N) + \sum_{i=k}^{N-1} f_i(x_i, \psi_{i+1}(x_{i+1}), \eta_i(x_i))\Delta_i\big)\mu^x(\mathrm{d}x_{k+1}, \ldots, \mathrm{d}x_N) \\
- y_k^R(x) : \quad \psi_i \in [\mathcal{S}_i^Y]_y, \quad \eta_{l,i} \in [\mathcal{S}_i^{Z,l}]_z\big\}
\end{aligned}$$

where μ^x is the law of $(X_{k+1}^{k,x}, \ldots, X_N^{k,x})$, and $[\mathcal{S}_k^Y]_y$ and $[\mathcal{S}_k^{Z,l}]_z$ are defined in (2.4.28). Notice that all functions in $\mathcal{G}'_k$ are bounded by $C_\Psi + C_y$. In view of the definition (2.4.12), $\bar{y}_k^{R,M} - y_k^R$ belongs to the set $\mathcal{G}'_k$. Since the expectation $\mathbb{E}_N^M[|\bar{y}_k^{R,M}(X_k) - y_k^R(X_k)|^2]$ integrates only w.r.t. the law of X_k, we clearly have

$$\begin{aligned}
\mathbb{P}(B_k^{Y,M}) &= \mathbb{P}\big(\varepsilon_{k,B}^Y + 2\mathbb{E}_N^M[|\bar{y}_k^{R,M}(X_k) - y_k^R(X_k)|^2] < \|\bar{y}_k^{R,M} - y_k^R\|_{k,M}^2\big) \\
&\leq \mathbb{P}\big(\exists G \in \mathcal{G}'_k : \varepsilon_{k,B}^Y + 2\mathbb{E}[G(X_k)^2] < \|G\|_{k,M}^2\big).
\end{aligned} \tag{2.4.29}$$

The latter probability is equal to $\mathbb{P}\big(\exists G \in \mathcal{G}'_k : \frac{\|G\|_{k,M}^2 - \mathbb{E}[G(X_k)^2]}{2\varepsilon_{k,B}^Y + \|G\|_{k,M}^2 + \mathbb{E}[G(X_k)^2]} > \frac{1}{3}\big)$ which can be estimated by applying Lemma 2.5.1 in 2.5.3: it gives

$$\mathbb{P}(B_k^{Y,M}) \leq 4\mathbb{E}\big(\mathcal{N}_1\big(\frac{2\varepsilon_{k,B}^Y}{15}, [\mathcal{G}'_k]^2, X_k^{1:M}\big)\big)\exp\big(-\frac{M\varepsilon_{k,B}^Y}{60(C_\Psi + C_y)^2}\big),$$

where we use the short notation $[\mathcal{G}'_k]^2 := \{g^2 : g \in \mathcal{G}'_k\}$. Thus, it remains to upper bound the $\mathbf{L}_1$ ε-covering number of $[\mathcal{G}'_k]^2$ and, for this, we exhibit a particular cover. Write $\mu_i^x(\mathrm{d}x_i)$ for the law of $X_i^{k,x}$; then define the probability measure

$$\nu_i^M(\mathrm{d}x_i) := \frac{1}{M}\sum_{m=1}^M \mu_i^{X_k^m}(\mathrm{d}x_i)$$

and denote by $[\mathcal{S}_i^Y(\varepsilon')]_y$ (resp. $[\mathcal{S}_i^{Z,l}(\varepsilon')]_z$) a $\mathbf{L}_1(\nu_i^M)$ ε'-cover of $[\mathcal{S}_i^Y]_y$ (resp $[\mathcal{S}_i^{Z,l}]_z$) where $\varepsilon' = \frac{\varepsilon}{4L_f T^{\frac{1+\theta_L}{2}}(q+1)(C_\Psi + C_y)}$. We claim that

$$\begin{aligned}[][\mathcal{G}'_k(\varepsilon)]^2 :=& \big\{x \in \mathbb{R}^d \mapsto \big[\int \big(\Phi(x_N) + \sum_{i=k}^{N-1} f_i(x_i, g_{i+1}^Y(x_{i+1}), g_i^Z(x_i))\Delta_i\big)\mu^x(\mathrm{d}x_{k+1},\dots,\mathrm{d}x_N) \\ &- y_k^R(x)\big]^2 : \quad g_i^Y \in [\mathcal{S}_i^Y(\varepsilon')]_y, \quad g_{l,i}^Z \in [\mathcal{S}_i^{Z,l}(\varepsilon')]_z\big\}\end{aligned}$$

ε-covers $[\mathcal{G}'_k]^2$ in the $\mathbf{L}_1$-norm w.r.t. the sample $X_k^{1:M}$. Indeed, for any $G \in \mathcal{G}'_k$ there exists $G_\varepsilon \in \mathcal{G}'_k(\varepsilon)$ such that

$$\begin{aligned}&\frac{1}{M}\sum_{m=1}^M |G(X_k^m)^2 - G_\varepsilon(X_k^m)^2| \\ &\le \frac{1}{M}\sum_{m=1}^M |G(X_k^m) - G_\varepsilon(X_k^m)| 2(C_\Psi + C_y) \\ &\le \frac{2}{M}\sum_{m=1}^M\sum_{i=k}^{N-1}\Delta_i \frac{(C_\Psi + C_y)L_f}{(T-t_i)^{\frac{1-\theta_L}{2}}} \\ &\quad\times \int \big\{|\psi_{i+1}(x_{i+1}) - g_{i+1}^Y(x_{i+1})| + |\eta_i(x_i) - g_i^Z(x_i)|\big\}\mu^{X_k^m}(\mathrm{d}x_{k+1},\dots,\mathrm{d}x_N) \\ &\le 2\sum_{i=k}^{N-1}\Delta_i\frac{(C_\Psi + C_y)L_f}{(T-t_i)^{\frac{1-\theta_L}{2}}}(q+1)\varepsilon' \le 2L_f\frac{T^{\frac{1+\theta_L}{2}}}{\frac{1+\theta_L}{2}}(q+1)(C_\Psi + C_y)\varepsilon' \le \varepsilon,\end{aligned}$$

where we have used $(\mathbf{A_F})$. Furthermore, following the method for the bound of $\mathbb{P}(A_k^{Y,M})$, we derive (for any $\varepsilon' \le C_y/2$)

$$\mathcal{N}_1(\varepsilon', [\mathcal{S}_i^Y]_y, \nu_i^M) \le 3\big(2e\frac{2C_y}{\varepsilon'}\log(3e\frac{2C_y}{\varepsilon'})\big)^{K_{0,i}+1} \le 3\big(\frac{22C_y}{5\varepsilon'}\big)^{2(K_{0,i}+1)} \tag{2.4.30}$$

using $2e\,x\log(3e\,x) \le (\frac{11x}{5})^2$ for $x \ge 4$. Similarly, $\mathcal{N}_1(\varepsilon', [\mathcal{S}_i^{Z,l}]_z, \nu_i^M) \le 3\big(\frac{22C_y}{5\varepsilon'\sqrt{\Delta_i}}\big)^{2(K_{l,i}+1)}$ for $\varepsilon' \le C_y/(2\sqrt{\Delta_i})$. Finally, we obtain

$$\begin{aligned}\mathcal{N}_1\big(\frac{2\varepsilon_{k,B}^Y}{15}, [\mathcal{G}'_k]^2, X_k^{1:M}\big) &\le \prod_{i=k}^{N-1} 3^{(q+1)}\big(\frac{22C_y}{5\varepsilon'}\big)^{2(K_{0,i}+1)}\big(\frac{22C_y}{5\varepsilon'\sqrt{\Delta_i}}\big)^{2\sum_{l=1}^q(K_{l,i}+1)} \\ &= \prod_{i=k}^{N-1} 3^{(q+1)}\Delta_i^{K_{0,i}+1}\big(\frac{132L_f T^{\frac{1+\theta_L}{2}}(q+1)(C_\Psi + C_y)C_y}{\varepsilon_{k,B}^Y\sqrt{\Delta_i}}\big)^{2\sum_{l=0}^q(K_{l,i}+1)}\end{aligned}$$

if $\varepsilon' = \frac{\varepsilon^Y_{k,B}}{30 L_f T^{\frac{1+\theta_L}{2}}(q+1)(C_\Psi + C_y)} \le \frac{C_y}{2} \wedge \frac{C_y}{2\sqrt{\Delta_k}} \wedge \cdots \wedge \frac{C_y}{2\sqrt{\Delta_{N-1}}}$. □

Bounds on $\mathbb{P}(C^{Y,M}_k)$ and $\mathbb{P}(C^{Z,M}_k)$ We detail the estimate only for $\mathbb{P}(C^{Y,M}_k)$. Define $\mathcal{G}''_k := \{x \mapsto g(x) - y^R_k(x) : g \in [\mathcal{S}^Y_k]_y\}$. The elements of $\mathcal{G}''_k$ are absolutely bounded by $2C_y$. Similarly to (2.4.29), observe that

$$
\begin{aligned}
\mathbb{P}(C^{Y,M}_k) &\le \mathbb{P}\big(\exists G \in \mathcal{G}''_k : \mathbb{E}[G(X_k)^2] > \frac{2}{M}\sum_{m=1}^M G(X^m_k)^2 + \varepsilon^Y_{k,C}\big) \\
&= \mathbb{P}\big(\exists G \in \mathcal{G}''_k : \frac{\mathbb{E}[G(X_k)^2] - \|G\|^2_{k,M}}{2\varepsilon^Y_{k,C} + \|G\|^2_{k,M} + \mathbb{E}[G(X_k)^2]} > \frac{1}{3}\big) \\
&\le 4\mathbb{E}\big(\mathcal{N}_1\big(\frac{\varepsilon^Y_{k,C}}{12}, [\mathcal{G}''_k]^2, X^{1:M}_k\big)\big) \exp\big(-\frac{M\varepsilon^Y_{k,C}}{507C_y^2}\big)
\end{aligned}
$$

where the second inequality follows from the second relation of Lemma 2.5.1 in 2.5.3. Observe that $\mathcal{N}_1\big(\frac{\varepsilon^Y_{k,C}}{12}, [\mathcal{G}''_k]^2, X^{1:M}_k\big) \le \mathcal{N}_1\big(\frac{\varepsilon^Y_{k,C}}{48C_y}, [\mathcal{S}^Y_k]_y, X^{1:M}_k\big)$. (2.4.30) is valid also for the empirical measure associated to $X^{1:M}_k$ and we obtain

$$
\mathbb{P}(C^{Y,M}_k) \le 12\big(\frac{1056C_y^2}{5\varepsilon^Y_{k,C}}\big)^{2(K_{0,k}+1)} \exp\big(-\frac{M\varepsilon^Y_{k,C}}{507C_y^2}\big)
$$

for $\frac{\varepsilon^Y_{k,C}}{48C_y} \le \frac{C_y}{2}$.

2.5 Appendix

2.5.1 SVD-optimal coefficients

Let us study the set of coefficients $\alpha \in \mathbb{R}^{K_{l,k}}$ minimizing $\|\alpha \cdot p_{l,k} - S\|^2_{k,M}$ using the Singular Value Decomposition (see [GVL96, Theorem 2.5.2 and Theorem 5.5.1]). The SVD of the $M \times K_{l,k}$-matrix $P_{l,k} = (p^{i,m}_{l,k})_{m,i}$ (with $M \ge K_{l,k}$) writes

$$
P_{l,k} = U P'_{l,k} V^\top \quad \text{with} \quad P'_{l,k} = \begin{pmatrix} \sigma_1 & & 0 \\ 0 & \ddots & \sigma_{K_{l,k}} \\ 0 & \cdots & 0 \end{pmatrix}
$$

where U and V are two unitary matrices respectively of size $M \times M$ and $K_{l,k} \times K_{l,k}$, and $\sigma_1 \ge \cdots \ge \sigma_{K_{l,k}} \ge 0$. If $P_{l,k}$ is a full rank matrix ($\sigma_{K_{l,k}} > 0$), the set of minimizers of $\|\alpha \cdot p_{l,k} - S\|^2_{k,M}$ reduces to a single element, whereas there are infinitely many minimizers for $\operatorname{rank}(P_{l,k}) < K_{l,k}$. Our choice of SVD-optimal solution consists of taking the element with minimal Euclidean norm which is given by

$$
\alpha^\star = V \begin{pmatrix} \cdots \\ \mathbf{1}_{\sigma_i > 0} \frac{(U^\top S)_i}{\sigma_i} \\ \cdots \end{pmatrix}. \tag{2.5.1}
$$

Proof. (of Proposition 2.4.1) The linearity property is clear from (2.5.1). The Pythagoras decomposition yields, for any $\alpha \in \mathbb{R}^M$, that $\|\alpha \cdot p_{l,k} - S\|^2_{k,M} = \|\alpha^\star \cdot p_{l,k} - S\|^2_{k,M} + \|(\alpha^\star - \alpha) \cdot p_{l,k}\|^2_{k,M}$. Taking $\alpha = 0$ gives statement ii). To prove iii), it is enough to use the linear relation (2.5.1) and to observe that $U, V, (\sigma_i)_i$ are $\mathcal{Q}$-measurable. $\square$

2.5.2 Proof of Proposition 2.3.9

We set $\mathcal{T}_R := \mathbb{E}([\mathcal{N} - (-R) \vee \mathcal{N} \wedge R]^2)$ where $\mathcal{N}$ is a Gaussian random variable with mean 0 and variance 1. An explicit computation gives

$$\mathcal{T}_R = 2\big(\mathbb{P}(\mathcal{N} > R)(R^2+1) - R\frac{e^{-\frac{1}{2}R^2}}{\sqrt{2\pi}}\big) \leq 2\mathbb{P}(\mathcal{N} > R)(R^2 + 1 - R^2) \leq 2e^{-\frac{1}{2}R^2},$$

where the two last inequalities are derived from the Mill inequality and the Markov exponential inequality.

Now, we follow the arguments of Lemma dfsd and we consider $\gamma \in (0, +\infty)^N$ such that $8q(\Delta_k + \frac{1}{\gamma_k})\frac{L_f^2}{(T-t_k)^{1-\theta_L}} \leq 1$ for $0 \leq k < N$. Define $\Delta Y_k := Y_k - Y_k^R$ and $\Delta Z_k := Z_k - Z_k^R$.
Preliminary bound on ΔZ. Applying the Cauchy-Schwartz inequality and the almost sure bounds on Y and Y^R (Proposition faf), we obtain:

$$\begin{aligned}
\Delta_k |\Delta Z_k|^2 &= \Delta_k^{-1}\big|\mathbb{E}_k\big[Y_{k+1}\Delta W_k - Y^R_{k+1}[\Delta W_k]_w\big]\big|^2 \\
&\leq 2\Delta_k^{-1}\big|\mathbb{E}_k[Y_{k+1}(\Delta W_k - [\Delta W_k]_w)]\big|^2 + 2\Delta_k^{-1}\big|\mathbb{E}_k\big[\Delta Y_{k+1}[\Delta W_k]_w\big]\big|^2 \\
&\leq 2qC_y^2\mathcal{T}_R + 2q\big(\mathbb{E}_k[\Delta Y^2_{k+1}] - (\mathbb{E}_k[\Delta Y_{k+1}])^2\big). \qquad (2.5.2)
\end{aligned}$$

Bound on ΔY. Using Young's inequality $(a+b)^2 \leq (1 + \Delta_k\gamma_k)a^2 + (1 + \frac{1}{\Delta_k\gamma_k})b^2$, the Lipschitz property of $(y, z) \mapsto f_k(y, z)$, and using (2.5.2), we obtain

$$\begin{aligned}
\Delta Y_k^2 \leq& (1 + \Delta_k\gamma_k)(\mathbb{E}_k[\Delta Y_{k+1}])^2 \\
&+ 2(\Delta_k + \frac{1}{\gamma_k})\frac{L_f^2}{(T-t_k)^{1-\theta_L}}\Delta_k(\mathbb{E}_k[\Delta Y^2_{k+1}] + |\Delta Z_k|^2) \qquad (2.5.3) \\
\leq& \big(1 + \Delta_k\gamma_k - 4q(\Delta_k + \frac{1}{\gamma_k})\frac{L_f^2}{(T-t_k)^{1-\theta_L}}\big)(\mathbb{E}_k[\Delta Y_{k+1}])^2 \qquad (2.5.4) \\
&+ 2(\Delta_k + \frac{1}{\gamma_k})\frac{L_f^2}{(T-t_k)^{1-\theta_L}}(\Delta_k + 2q)\mathbb{E}_k[\Delta Y^2_{k+1}] \\
&+ 4qC_y^2(\Delta_k + \frac{1}{\gamma_k})\frac{L_f^2}{(T-t_k)^{1-\theta_L}}\mathcal{T}_R.
\end{aligned}$$

The condition $8q(\Delta_k + \frac{1}{\gamma_k})\frac{L_f^2}{(T-t_k)^{1-\theta_L}} \leq 1$ ensures that $1 + \Delta_k\gamma_k - 4q(\Delta_k + \frac{1}{\gamma_k})\frac{L_f^2}{(T-t_k)^{1-\theta_L}} \geq 0$;

this given, we may use Jensen's inequality on the term (2.5.4) to obtain:

$$\begin{aligned}
\Delta Y_k^2 &\leq \big(1+\Delta_k\gamma_k + 2(\Delta_k+\frac{1}{\gamma_k})\frac{L_f^2}{(T-t_k)^{1-\theta_L}}\Delta_k\big)\mathbb{E}_k[\Delta Y_{k+1}^2] \\
&\quad + 4qC_y^2(\Delta_k+\frac{1}{\gamma_k})\frac{L_f^2}{(T-t_k)^{1-\theta_L}}\mathcal{T}_R \\
&\leq \big(1+\Delta_k\gamma_k+\frac{\Delta_k}{4}\big)\mathbb{E}_k[\Delta Y_{k+1}^2] + \frac{1}{2}C_y^2\mathcal{T}_R
\end{aligned}$$

using again the relation between Δ_k and $1/\gamma_k$. Multiplying by $\lambda_k := \prod_{j=0}^{k-1}(1+\Delta_j\gamma_j+\frac{\Delta_j}{4})$, taking conditional expectation $\mathbb{E}_i$, and summing over $k=i,\ldots,N-1$, we obtain a pointwise uniform bound for ΔY_i^2:

$$\Delta Y_i^2\Gamma_i \leq \Delta Y_i^2\lambda_i \leq \frac{1}{2}C_y^2 e^{T/4}\Gamma_N N\mathcal{T}_R. \tag{2.5.5}$$

Final bound on ΔZ. (2.5.2) yields:

$$\begin{aligned}
\sum_{k=i}^{N-1}\Gamma_k\mathbb{E}_i[|\Delta Z_k|^2]\Delta_k &\leq 2qC_y^2\Gamma_N N\mathcal{T}_R + 2q\sum_{k=i}^{N-1}\big(\mathbb{E}_i[\Delta Y_{k+1}^2]-\mathbb{E}_i(\mathbb{E}_k[\Delta Y_{k+1}])^2\big)\Gamma_{k+1} \\
&\leq 2qC_y^2\Gamma_N N\mathcal{T}_R + 2q\sum_{k=i}^{N-1}\big(\mathbb{E}_i[\Delta Y_k^2]-(1+\Delta_k\gamma_k)\mathbb{E}_i(\mathbb{E}_k[\Delta Y_{k+1}])^2\big)\Gamma_k.
\end{aligned}$$

Substituting the inequality (2.5.3), we obtain

$$\begin{aligned}
&\sum_{k=i}^{N-1}\Gamma_k\mathbb{E}_i[|\Delta Z_k|^2]\Delta_k \\
&\leq 2qC_y^2\Gamma_N N\mathcal{T}_R + 4q\sum_{k=i}^{N-1}(\Delta_k+\frac{1}{\gamma_k})\frac{L_f^2}{(T-t_k)^{1-\theta_L}}\Delta_k\Gamma_k(\mathbb{E}_i[\Delta Y_k^2]+\mathbb{E}_i[|\Delta Z_k|^2]) \\
&\leq 2qC_y^2\Gamma_N N\mathcal{T}_R + \frac{1}{2}\sum_{k=i}^{N-1}\Delta_k\Gamma_k(\mathbb{E}_i[\Delta Y_k^2]+\mathbb{E}_i[|\Delta Z_k|^2])
\end{aligned}$$

taking into account the relation between π and γ. Thus, we have

$$\sum_{k=i}^{N-1}\Gamma_k\mathbb{E}_i[|\Delta Z_k|^2]\Delta_k \leq C_y^2\Gamma_N N\mathcal{T}_R\big(4q+\frac{T}{2}\exp(\frac{T}{4})\big). \tag{2.5.6}$$

Observe that $\gamma_k := \frac{24qL_f^2}{(T-t_k)^{1-\theta_L}}$ defines an admissible choice, provided that $C_\pi L_f^2 \leq \frac{1}{12q}$. It gives $1 \leq \Gamma_i \leq \Gamma_N \leq \exp\big(\frac{24qL_f^2}{\theta_L}T^{\theta_L}\big)$. Plugging this estimate into (2.5.5) and (2.5.6), and using the bound on $\mathcal{T}_R$, we obtain the final result. □

2.5.3 Upper bound of a deviation probability, uniform over a class of functions

For the definition of the covering number $\mathcal{N}_1(\ldots)$ used below, we refer to the notation of Subsection 2.4.5.

Lemma 2.5.1. *Let $\mathcal{G}$ be a countable set of functions $g:\mathbb{R}^d \mapsto [0,B]$ with $B>0$. Let $X, X^1,\ldots,X^M$*

$(M \geq 1)$ be i.i.d. $\mathbb{R}^d$ valued random variables. For any $\alpha > 0$ and $\varepsilon \in (0,1)$ one has

$$\mathbb{P}\big(\sup_{g\in\mathcal{G}} \frac{\frac{1}{M}\sum_{m=1}^M g(X^m) - \mathbb{E}[g(X)]}{\alpha + \frac{1}{M}\sum_{m=1}^M g(X^m) + \mathbb{E}[g(X)]} > \varepsilon\big) \leq 4\mathbb{E}\big(\mathcal{N}_1\big(\frac{\alpha\varepsilon}{5}, \mathcal{G}, X^{1:M}\big)\big) \exp\big(-\frac{3\varepsilon^2\alpha M}{40B}\big),$$

$$\mathbb{P}\big(\sup_{g\in\mathcal{G}} \frac{\mathbb{E}[g(X)] - \frac{1}{M}\sum_{m=1}^M g(X^m)}{\alpha + \frac{1}{M}\sum_{m=1}^M g(X^m) + \mathbb{E}[g(X)]} > \varepsilon\big) \leq 4\mathbb{E}\big(\mathcal{N}_1\big(\frac{\alpha\varepsilon}{8}, \mathcal{G}, X^{1:M}\big)\big) \exp\big(-\frac{6\varepsilon^2\alpha M}{169B}\big).$$

Proof. The first inequality is stated in [GKKW02, Theorem 11.6] for $B \geq 1$. For $B \in (0,1)$, we rescale the class of functions $\{g/B : g \in \mathcal{G}\}$ (now bounded by 1), replace α by α/B and apply the previous case: this gives the announced upper bound.
To establish the second inequality, we adapt the proof of the first inequality from the proof of [GKKW02, Theorem 11.6]. The first step consists in taking a ghost sample $\tilde{X}^{1:M}$ and observing that for a given $g \in \mathcal{G}$, $\mathbb{E}[g(X)] - \frac{1}{M}\sum_{m=1}^M g(X^m) > \varepsilon\big(\alpha + \frac{1}{M}\sum_{m=1}^M g(X^m) + \mathbb{E}[g(X)]\big)$ and $\mathbb{E}[g(X)] - \frac{1}{M}\sum_{m=1}^M g(\tilde{X}^m) \leq \frac{\varepsilon}{4}\big(\alpha + \frac{1}{M}\sum_{m=1}^M g(\tilde{X}^m) + \mathbb{E}[g(X)]\big)$ imply

$$\begin{aligned}(1+\frac{5\varepsilon}{8})\big(\frac{1}{M}\sum_{m=1}^M g(\tilde{X}^m) - \frac{1}{M}\sum_{m=1}^M g(X^m)\big)&\\ > \frac{3\varepsilon}{8}\big(2\alpha + \frac{1}{M}\sum_{m=1}^M g(X^m) + \frac{1}{M}\sum_{m=1}^M g(\tilde{X}^m)\big) + \frac{3\varepsilon}{4}\mathbb{E}[g(X)].&\end{aligned}$$

Since the r.h.s. positive, the l.h.s. is also positive; using $\frac{13}{8} \geq 1 + \frac{5\varepsilon}{8}$ implies

$$\frac{1}{M}\sum_{m=1}^M g(\tilde{X}^m) - \frac{1}{M}\sum_{m=1}^M g(X^m) > \frac{3\varepsilon}{13}\big(2\alpha + \frac{1}{M}\sum_{m=1}^M g(X^m) + \frac{1}{M}\sum_{m=1}^M g(\tilde{X}^m)\big).$$

Then we proceed as in [GKKW02, pp. 205-207] to show that the probability to estimate is bounded by

$$2\mathbb{P}\Big(\exists g \in \mathcal{G} : \frac{1}{M}\sum_{m=1}^M g(\tilde{X}^m) - \frac{1}{M}\sum_{m=1}^M g(X^m) > \frac{3\varepsilon}{13}\big(2\alpha + \frac{1}{M}\sum_{m=1}^M g(X^m) + \frac{1}{M}\sum_{m=1}^M g(\tilde{X}^m)\big)\Big)$$

for $M > \frac{8B}{\varepsilon^2\alpha}$ (however for $M \leq \frac{8B}{\varepsilon^2\alpha}$ the upper bound in Lemma 2.5.1 is obviously true). The rest of the proof is identical to [GKKW02, pp. 208-210], except that one should take a $\mathbf{L}_1$ δ-cover of $\mathcal{G}$ w.r.t. $X^{1:M}$ with $\delta = \frac{\alpha\varepsilon}{8}$ (instead of $\delta = \frac{\alpha\varepsilon}{5}$). It leads to a new upper bound, $4\mathbb{E}\big(\mathcal{N}_1\big(\frac{\alpha\varepsilon}{8}, \mathcal{G}, X^{1:M}\big)\big) \exp\big(-\frac{6\varepsilon^2\alpha M}{169B}\big)$. $\square$

3 A multilevel algorithm for the approximation of BSDEs

3.1 Introduction

Let $(\Omega^{(W)}, \mathcal{F}^{(W)}, (\mathcal{F}_t^{(W)}), \mathbb{P}^{(W)})$ be a filtered probability space generated by a q-dimensional Brownian Motion $(W_t)_{t\leq T}$ with the usual conditions, and $T > 0$ a finite terminal time. The purpose of this chapter is to design and analyze a numerical scheme to approximate the system of BSDEs

$$\left.\begin{array}{rcl} y_t &=& \Phi(X_T) - \int_t^T z_s dW_s, \\ \bar{y}_t &=& \int_t^T f(r, X_r, y_r + \bar{y}_r, z_r + \bar{z}_r)dr - \int_t^T \bar{z}_r dW_r, \end{array}\right\} \tag{3.1.1}$$

where X is the d-dimensional SDE

$$X_t = x_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s. \tag{3.1.2}$$

The linear BSDE (y, z) in (3.1.1) is approximated using a multilevel approach. Let $\big(\pi^{(k)} = \{0 = t_0^{(k)} < \ldots < t_{N^k}^{(k)} = T\}\big)_{k=0,\ldots,\kappa}$ be a refining sequence (i.e., $\pi^{(k-1)} \subset \pi^{(k)}$) of time-grids on $[0, T]$, and $X^{(k)}$ an appropriate discretization of the SDE X on $\pi^{(k)}$. Assuming that a discretization $(y^{(k-1)}, z^{(k-1)})$ of (y, z) on $\pi^{(k-1)}$ is available, define the discretization $(y^{(k)}, z^{(k)})$ on $\pi^{(k)}$ by defined by

$$\left.\begin{array}{rcl} y_{N^k}^{(k)} &:=& \Phi(X_{N^k}^{(k)}), \quad y_i^{(k)} := \mathbb{E}^{(W)}[\Phi(X_{N^k}^{(k)})|\mathcal{F}_{t_i^{(k)}}^{(W)}], \\ \Delta^{(k)} z_i^{(k)} &:=& \mathbb{E}^{(W)}[(\Phi(X_{N^k}^{(k)}) - y_{\alpha(i)}^{(k-1)} - \sum_{j=\alpha(i)+1}^{N^{k-1}-1} z_j^{(k-1)} \Delta W_j^{(k-1)})\Delta W_i^{(k)}|\mathcal{F}_{t_i^{(k)}}^{(W)}] \end{array}\right\} \tag{3.1.3}$$

where $\Delta^{(k)} = (t_{i+1}^{(k)} - t_i^{(k)})$, $\Delta W_i^{(k)} := W_{t_{i+1}^{(k)}} - W_{t_i^{(k)}}$ and $\Delta W_j^{(k-1)} := W_{t_{j+1}^{(k-1)}} - W_{t_j^{(k-1)}}$, and $\alpha(i)$ is the largest integer such that $t_{\alpha(i)}^{(k-1)} \leq t_i^{(k)}$. Equation (3.1.3) gives us the iterative structure of the multilevel algorithm: taking $\pi^{(0)}$ to be the trivial time grid $\{0, T\}$, one initializes the algorithm by $y_0^{(0)} = \mathbb{E}^{(W)}[\Phi(X_1^{(0)})]$ and $z_0^{(0)} = \mathbb{E}^{(W)}[\Phi(X_1^{(0)})]/T$, and then computes $(y^{(k)}, z^{(k)})$ for every $k > 0$ using (3.1.3). However, the conditional expectations $\mathbb{E}^{(W)}[\cdot|\mathcal{F}_{t_i^{(k)}}^{(W)}]$ cannot, in general, be computed explicitly. In order to obtain a fully implementable scheme, we replace the conditional expectation operator by Monte Carlo least-square regression. The resulting discrete time approximations $(y^{(k,M)}, z^{(k,M)})$ depend also on the simulations used for the Monte Carlo scheme. Observe that, since $(y_j^{(k-1)}, z_j^{(k-1)})$ is $\mathcal{F}_{t_j^{(k-1)}}^{(W)}$-measurable, the Tower Law for conditional expectations implies that the $(y^{(k-1)}, z^{(k-1)})$ terms in equation (3.1.3) vanish, and we are reduced to the MDP scheme $y_i^{(k)} = \mathbb{E}^{(W)}[\Phi(X_{N^k}^{(k)})|\mathcal{F}_{t_i^{(k)}}^{(W)}]$ and $z_i^{(k)} := \mathbb{E}^{(W)}[\Phi(X_{N^k}^{(k)})\Delta W_i^{(k)}|\mathcal{F}_{t_i^{(k)}}^{(W)}]/\Delta^{(k)}$ of Chapter 2. This reduction does not, in general, occur when the conditional expectation operator is replaced by Monte Carlo least-squares regression. We use independent clouds of simulations for each level k of the algorithm. The number of simulations is allowed to differ from level to level. We compute explicit upper bounds for the error of the multilevel scheme $\mathbb{E}[|y_i^{(k)} - y_i^{(k,M)}|^2]$ and $\mathbb{E}[|z_i^{(k)} - z_i^{(k,M)}|^2]$ for each $k \in \{0, \ldots, \kappa\}$ and $i \in \{0, \ldots, N^k - 1\}$ in terms of the choice of basis functions for the least-square regression, the number of simulations, and the error of the approximation on level $k-1$ (Theorem 3.3.9). The impact of the multilevel structure in (3.1.3) is that the approximations $(y^{(k-1,M)}, z^{(k-1,M)})$ act as control variates for the least-squares regression, reducing the variance of the least-squares regression at level k. The upper bounds of error are then used to calibrate the basis functions and sample sizes sequentially from level κ to level 0 so that the overall global error at level κ

$$\max_{0\leq i\leq N-1}\mathbb{E}[|y_i^{(\kappa)}-y_i^{(\kappa,M)}|^2]+\sum_{i=0}^{N_\kappa-1}\mathbb{E}[|z_i^{(\kappa)}-z_i^{(\kappa,M)}|^2]\Delta^{(\kappa)}$$

satisfies an error tolerance $\epsilon=O(N^{-1})$. We use a local polynomial basis; the polynomials are of degree n, and are localized on hypercubes. By carefully selecting the number simulations used for each level of the scheme, we are able to show that complexity of the multilevel scheme may be up to order 1 less than that of the LSMDP algorithm of Chapter 2. We give a specific example in the case where the terminal condition $\Phi(\cdot)$ is n-times differentiable: the complexity of the multilevel scheme is $O\big(\epsilon^{-2-\frac{d}{2(n-1)}}\log^{1+d}(\epsilon^{-1})\big)$ (Proposition 3.3.11), whereas that of LSMDP is $O\big(\epsilon^{-3-\frac{d}{2(n-1)}}\log^{1+d}(\epsilon^{-1})\big)$ (Proposition 3.3.13). On the other hand, if the terminal condition satisfies only Lipschitz continuity, the complexity of the two algorithms appears to be the same (Proposition 3.3.10 versus Proposition 3.3.12), at least in terms of this analysis based on the upper bounds of the error and the use of local polynomial basis. In this case, the complexity of the multilevel algorithm is $O\big(\epsilon^{-\frac{d}{2(n-1)}-\frac{d}{2}-2}\log^{d+2}(\epsilon^{-1})\big)$. The main factor that differentiates between the complexity of the LSMDP scheme and multilevel scheme, at least in terms of the analysis with the maximum error bound, is the number of basis functions used for the least-squares regressions: the complexity of the multilevel scheme is dependent on the largest number of basis functions used, whereas LSMDP depends on the total number of basis functions used for the time grid. Therefore, it would be instructive to perform this comparison when other basis functions are used, but this is outside the scope of this work.

The nonlinear BSDE $(\bar{y},\bar{z})$ in (3.1.1) is computed using an adaptation of the LSMDP scheme of Chapter 2. A detailed error analysis of this algorithm is provided; as for LSMDP, a priori estimates for discrete BSDEs are the main tool used in the analysis. As for the multilevel scheme, the upper bounds of the error are used to calibrate the algorithm. The main difference between the scheme introduced in this chapter and LSMDP is that, in this chapter, the scheme uses an independent cloud of simulations for the least-squares regressions on each time-point, whereas LSMDP uses a single cloud of simulations for every time-point. This enables the algorithm to avoid complicated error terms due to interdependence of the regression coefficients, and this seems to reduce the overall complexity of the algorithm. Using a local polynomial bases with polynomials of degree n for the least-squares regressions, the complexity of this algorithm (Proposition 3.4.21) is $O\big((\epsilon^{-1-d/2}+\epsilon^{-4-\frac{d}{2(n-1)}}\mathbf{1}_{d<6(n-1)/(n-2)})\log^{1+d}(\epsilon^{-1})\big)$. This implies that the complexity of the complete algorithm approximating the system of BSDEs (3.1.1) is dominated by the complexity of the multilevel scheme, except in the special case of $n=2$ and $d\leq 3$, so the overall complexity is $O\big(\epsilon^{-\frac{d}{2(n-1)}-\frac{d}{2}-2}\log^{d+2}(\epsilon^{-1})\big)$.

In fact, $(y+\bar{y},z+\bar{z})$, the sum of the BSDEs in (3.1.1), is equal to (Y,Z), the BSDE solving

$$Y_t=\Phi(X_T)+\int_t^T f(r,X_r,Y_r,Z_r)dr-\int_t^T Z_r dW_r. \tag{3.1.4}$$

The approximation of (Y, Z) is the main application of the algorithm described above. It is also possible to approximate (Y, Z) directly using the adapted LSMDP, without the decomposition (3.1.1). In order to provide a fair comparison between the two methods, we compute upper bounds of the error of the approximation of the full BSDE (3.1.4) using adaped LSMDP, and use these bounds to calibrate the algorithm and compute the resulting complexity. The resulting complexity (Proposition 3.5.2) of adapted LSMDP applied to (Y, Z) is $O\big((\epsilon^{-3-\frac{d}{2}-\frac{d}{2(n-1)}}\mathbf{1}_{d\geq 2} + C\epsilon^{-4-\frac{d}{2(n-1)}}\mathbf{1}_{d=1})\log^{1+d}(\epsilon^{-1})\big)$, which is order 1 worse that the complexity of the algorithm for the the computation of the system of BSDEs (3.1.1) even when the multilevel algorithm is replaced by LSMDP. As stated in the above discussion, the complexity of the algorithm for the system of BSDEs may be even lower when multilevel is in use. The complexity of the adapted LSMDP applied to the $(\bar{y}, \bar{z})$ in the system of BSDEs (3.1.1) is lower than the complexity of the adapted LSMDP applied to (Y, Z) (1.1.1). The main reason for this is that the almost sure bounds of $|\bar{y}_i|$ and $|\bar{z}_i|$ converge to zero as one approaches the terminal time. This is because the BSDE $(\bar{y}, \bar{z})$ has zero terminal condition. In contrast, the almost sure bounds of $|Y_i|$ and $|Z_i|$ are constant because the terminal condition is not zero. The impact of this is that the adapted LSMDP algorithm applied to $(\bar{y}, \bar{z})$ suffers from less variance due to least-squares regression. We also analyze the application of the adapted LSMDP to (Y, Z) in the case where the terminal condition Φ is n-times continuously differentiable with bounded derivatives in order to compare it to the LSMDP of Chapter 2. The complexity of the adapted LSMDP is $O\big(\epsilon^{-4-\frac{d}{2(n-1)}}\log^{1+d}(N)\big)$ (Proposition 3.5.3), whereas the complexity of LSMDP is $O\big(\epsilon^{-4-\frac{d}{(n-1)}}\log^{2+d}(N)\big)$ (Section 2.4.3), so the adapted LSMDP has a higher efficiency. This is due to the removal of the complicated interdependence between the regression coefficients.

The use of the decomposition (3.1.1) is not original. For example, a numerical study can be found in [BS12]. However, this chapter contains, to the best of our knowledge, the only explicit theoretical comparison between approximation of the system of BSDEs (3.1.1) and the approximation of the full BSDE (3.1.4) with the use of Monte Carlo least-squares regression schemes.

We would also like to mention the idea of using a martingale basis for computing the BSDEs (y, z) [BS12]. The idea is to select basis functions whose conditional expectations are a priori known. This implies that one avoids the use of Monte Carlo least-squares to approximate (y, z), thereby reducing the cost of numerical scheme significantly. However, the existence of a martingale basis may not be guaranteed for general forwards process X. The multilevel scheme in this paper gives a efficient and generic method for approximating the BSDE (y, z), which is available when there is no known martingale basis.

The multilevel algorithm given in this chapter bears some conceptual similarity to the multilevel scheme for SDEs of [Gil08]: we use approximations on coarser time-grids as control variates for the approximations on the refined grids. The structure of the multilevel recursion (3.1.3) and the error analysis are based on BSDE techniques and bear little resemblance to the SDE techniques. The use of multilevel for the approximation of BSDEs is, to the best of our knowledge, novel.

The chapter is structured as follows. The notation used throughout the article is given in Section 3.1.1, and the standing assumptions in Section 3.1.2. The discretization of the SDE X (3.1.2) is given in Section 3.1.3. Some basic results on the linear BSDE (y, z) that will be used throughout this chapter are given in Section 3.2. The multilevel algorithm is given in Section 3.3. It is constructed in Section 3.3.1. The main results on the error analysis are contained in Section 3.3.2. The calibration of the multilevel algorithm and computation of the complexity are contained in Section 3.3.3. The adapted LSMDP algorithm for the nonlinear BSDE $(\bar{y}, \bar{z})$ is given in Section 3.4; the error analysis of this scheme is in Section 3.4.4, and the calibration of the algorithm and it's complexity are found in Section 3.4.5. The adapted LSMDP algorithm for the full BSDE (Y, Z) is given in Section 3.5.

3.1.1 Notation

For a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and sub-σ-algebra $\mathcal{G}$, we write $\mathbf{L}_2(\mathcal{G})$ for the space of $\mathcal{H}$-measurable, square integrable random variables.

We write $\mathbf{L}_2([0,T] \times \Omega^{(W)})$ for the space of $\mathcal{B}([0,T]) \otimes \mathcal{F}^{(W)}$-measurable processes that are square integrable with respect to $dt \times \mathbb{P}$.

We write $\mathcal{P}$ for the predictable σ-algebra, and $\mathcal{H}^2$ for the space of $\mathcal{P}$-measurable process that are square integrable with respect to $dt \times \mathbb{P}$.

For any vector or matrix V, we denote it's transpose by $V^\top$.

We introduce the following notation to deal with the sequence of refining time-grids and discrete processes.

- For each $k \geq 0$, we define by

$$\pi^{(k)} = \left\{0 = t_0^{(k)} < \ldots < t_{2^k}^{(k)} = T \;:\; t_i^{(k)} = Ti2^{-(k)},\ i \in \{0, \ldots, 2^k\}\right\} \tag{3.1.5}$$

  the uniform time-grid on $[0,T]$ with $2^k + 1$ points. The time increments are denoted by $\Delta^{(k)} := \frac{T}{2^k}$. We denote by $\Delta W_i^{(k)} := W_{t_{i+1}^{(k)}} - W_{t_i^{(k)}}$ the increments of the Brownian Motion on the time-grid $\pi^{(k)}$.

- $\pi^{(k+1)}$ is refinement of $\pi^{(k)}$. To deal with the overlapping time-points, we introduce the function $\alpha : \{0, \ldots, 2^k\} \to \{0, \ldots, 2^{k-1}\}$ given by

$$\alpha(i) = \begin{cases} \frac{i}{2} & \text{if } i \text{ is even}, \\ \frac{i-1}{2} & \text{if } i \text{ is odd}. \end{cases} \tag{3.1.6}$$

  It's j-th iteration is denoted by $\alpha_j : \{0, \ldots, 2^k\} \to \{0, \ldots, 2^{k-j}\}$, and $\alpha_j(i) := (\alpha)^j(i)$. Although α and α_j depend on k, we omit this dependence in the notation, since it will be obvious from the context which k is meant.

- We define the σ-algebras $\mathcal{F}_i^{(k)} := \mathcal{F}_{t_i^{(k)}}^{(W)}$ and denote the conditional expectation associated to this σ-algebra by $\mathbb{E}_i^k[\cdot] := \mathbb{E}[\cdot|\mathcal{F}_i^{(k)}]$. The sequence $(\mathcal{F}_i^{(k)}) = (\mathcal{F}_i^{(k)})_{0 \leq i \leq N}$ forms a filtration in discrete time.

- We say that a stochastic process X is discrete on $\pi^{(k)}$ if it is piecewise constant with nodes at the time points of $\pi^{(k)}$; moreover, we say that X is $(\mathcal{F}_i^{(k)})$-adapted if, for each $i \in \{0, \ldots, N\}$,

X_i is $\mathcal{F}_i^{(k)}$-measurable, and that it is an $(\mathcal{F}_i^{(k)})$-martingale if it is a martingale with respect to the filtration $(\mathcal{F}_i^{(k)})$.

- In Section 3.3.3, we will furthermore denote the number of time-points in the finest time-grid in use by N.

3.1.2 Assumptions

The following assumptions will hold throughout the entirety of this chapter.

$(\mathbf{A_{b,\sigma}})$ The coefficients of the SDE (3.1.2) satisfy the following properties.

(i) $(t,x) \in [0,T]\times\mathbb{R}^d \mapsto b(t,x)$ is $\mathbb{R}^d$-valued, measurable and uniformly bounded. Moreover, $b(t,\cdot)$ is twice continuously differentiable with uniformly bounded derivatives and Hölder continuous second derivative, and $b(\cdot,x)$ is 1/2-Hölder continuous.

(ii) $(t,x) \in [0,T]\times\mathbb{R}^d \mapsto \sigma(t,x)$ is $\mathbb{R}^d\times\mathbb{R}^q$-valued, measurable and uniformly bounded. Moreover, $\sigma(t,\cdot)$ is twice continuously differentiable with uniformly bounded derivatives and Hölder continuous second derivative, and $\sigma(\cdot,x)$ is 1/2-Hölder continuous.

$(\mathbf{A}_{\text{u.e.}})$ $\sigma(\cdot)$ satisfies a uniformly elliptic condition: there exists some finite $\beta > 0$ such that, for any $\zeta \in \mathbb{R}^d$, $\zeta^\top\sigma(t,x)\sigma(t,x)^\top\zeta \geq \beta|\zeta|^2$ for all $(t,x) \in [0,T]\times\mathbb{R}^d$.

$(\mathbf{A_\Phi})$ The terminal condition $\Phi : \mathbb{R}^d \to \mathbb{R}$ of the BSDE (y,z) in (3.1.1) is measurable and satisfies the following properties.

i) Φ is Lipschitz continuous with Lipschitz constant L_Φ.

ii) Φ is uniformly bounded by C_Φ.

$(\mathbf{A_f})$ The driver $f : [0,T]\times\mathbb{R}^d\times\mathbb{R}\times\mathbb{R}^q \to \mathbb{R}$ of the BSDE $(\bar{y},\bar{z})$ in (3.1.1) is measurable and satisfies the following properties.

i) $f(t,\cdot)$ is Lipschitz continuous with uniform, finite Lipschitz constant $L_f > 0$: for all $t \in [0,T)$, $(x,y,z),(x',y',z') \in \mathbb{R}^d\times\mathbb{R}\times\mathbb{R}^q$,

$$|f(t,x,y,z) - f(t,x',y',z')| \leq L_f\{|x-x'| + |y-y'| + |z-z'|\}$$

ii) $f(t,x,0,0)$ is uniformly bounded by a finite constant $C_f > 0$: for all $(t,x) \in [0,T)\times\mathbb{R}^d$, $|f(t,x,0,0)| \leq C_f$.

Definition 3.1.1 (The constant C)**.** *Throughout this chapter, we shall use the convention that $C \geq 0$ is a deterministic value that can depend only on the terminal time T, the dimensions d, q, the uniform bounds $\|b\|_\infty$, $\|\nabla_x b\|_\infty$, $\|\nabla_x^2 b\|_\infty$, $\|\sigma\|_\infty$, $\|\nabla_x\sigma\|_\infty$, $\|\nabla_x^2\sigma\|_\infty$, C_Φ, C_f, the uniform elliptic parameter β, and the Lipschitz constants L_Φ, L_f.*

To simplify the notation, C is allowed to change it's value from line to line.

Remark. It is very important that the constant C given in Definition 3.1.1 does not depend on the time-grids $\{\pi^{(k)} : k = 0,\dots,\kappa\}$, on the level k or on the total number of levels κ; it also does not depend on the number of time points in the time-grids.

3.1.3 The forward process

For each $t \in [0, T)$ and $x \in \mathbb{R}^d$, we define the Markovian SDEs

$$X_s^{(t,x)} = x \text{ for } s < t,$$
$$X_s^{(t,x)} = x + \int_t^s b(r, X_r^{(t,x)})dr + \int_t^s \sigma(r, X_r^{(t,x)})dW_r \text{ for } s \geq t. \tag{3.1.7}$$

Under $(\mathbf{A}_{\mathbf{b},\sigma})$, (3.1.7) has a unique strong solution for each $(t, x) \in [0, T) \times \mathbb{R}^d$ [RY99, Theorem IX.2.1]. The diffusion X in (3.1.4) is equal to $X^{(0,x_0)}$. Moreover, the proof of [RY99, Theorem IX.2.4] gives us the following continuity properties.

Lemma 3.1.2. *For all $t \in [0, T)$, $r, s \in [t, T]$, and $x_1, x_2 \in \mathbb{R}^d$,*

$$\mathbb{E}^W[|X_s^{(t,x_1)} - X_s^{(t,x_2)}|^2|\mathcal{F}_t^{(W)}] \leq C|x_1 - x_2|^2, \quad \mathbb{E}^W[|X_s^{(t,x_1)} - X_r^{(t,x_1)}|^2|\mathcal{F}_t^{(W)}] \leq C|s - r|.$$

It is also well know that $X^{(t,x)}$ is a time-inhomogeneous Markov process [RY99, Theorem IX.1.9].

Assume that we cannot explicitly compute the marginals X_t of the diffusion X. For each time-grid $\pi^{(k)}$, we will approximate the diffusion X by a $\pi^{(k)}$-discrete process $X^{(k)}$. Let $i \in \{0\ldots, 2^k\}$ and $x \in \mathbb{R}^d$, and let $(X_s^{(t_i^{(k)},x)})_{s \geq t_i^{(k)}}$ be the Markov process given by (3.1.7) with $t = t_i^{(k)}$. We assume that there is a $\pi^{(k)}$-Markov chain $(X_j^{(k,i,x)})_{i \leq j \leq N}$ (with $X_i^{(k,i,x)} = x$) satisfying the property

$$\mathbb{E}_i^k[|X_{t_j^{(k)}}^{(t_i^{(k)},x)} - X_j^{(k,i,x)}|^p] \leq C(\Delta^{(k)})^{p/2} \quad \text{for all } p \in \{2, 3, 4\}, \quad \text{for all } j \in \{i, \ldots 2^k\}. \tag{3.1.8}$$

Finally, we assume there is a function $f_{j,r}^{(k)} : \Omega^{(W)} \times \mathbb{R}^d \to \mathbb{R}^d$ which is $\sigma(W_t - W_{t_i^{(k)}} : t \geq t_i^{(k)}) \otimes \mathcal{B}(\mathbb{R}^d)$-measurable such that $X_r^{(k,i,x)} = f_{j,r}^{(k)}(X_j^{(k,i,x)})$ for all $r \geq j \geq i$ and $x \in \mathbb{R}^d$.

We shall use the Markov chain $X^{(k)} := X^{(k,0,x_0)}$ to approximate the diffusion X on the time-grid $\pi^{(k)}$.

To give an example of an approximation scheme, the Markov chain $X^{(k,i,x)}$ generated by the Euler scheme

$$X_j^{(k,i,x)} = x + \sum_{r=i}^{j-1}\{b(t_r^{(k)}, X_r^{(k,i,x)})\Delta^{(k)} + \sigma(t_r^{(k)}, X_r^{(k,i,x)})\Delta W_r^{(k)}\} \tag{3.1.9}$$

with k sufficiently large satisfies the first inequality in (3.1.8) [Avi09, Theorem 5.3]. We can sequentially construct the deterministic function $F_{i,j}^{(k)} : \mathbb{R}^d \times (\mathbb{R}^q)^{j-i} \mapsto \mathbb{R}^d$ such that $X_j^{(k)} = F_{i,j}^{(k)}(X_i^{(k)}, \Delta W_i^{(k)}, \ldots,$
$\Delta W_{j-1}^{(k)})$, so $f_{i,j}^{(k)}(x) = F_{i,j}^{(k)}(x, \Delta W_i^{(k)}, \ldots, \Delta W_{j-1}^{(k)})$.

Remark. The first inequality in (3.1.8) combined with $(\mathbf{A}_\Phi)$ implies that

$$\mathbb{E}_i^k[|\Phi(X_T^{(t_i^{(k)},x)}) - \Phi(X_{2^k}^{(k,i,x)})|^p] \leq C(\Delta^{(k)})^{p/2} \quad \text{for all } p \in \{2, 3, 4\}.$$

A weaker bound is obtained in the case where Φ has only bounded variation [Avi09, Theorem 5.4].

Since we are now considering multiple time-grids - $\pi^{(0)}$ through $\pi^{(\kappa)}$ - we will need to define a Markov chain $(X_i^{(k)})_{0\leq i\leq 2^k}$ for each $k \in \{0,\dots,\kappa\}$. We require that (3.1.8) is satisfied by $X^{(k,i,x)}$ for all k. This raises problems in, for example, the case where $X^{(k,i,x)}$ is generated by the Euler scheme (3.1.9), where the condition in (3.1.8) is only satisfied on a sufficiently fine time-grid. Moreover, for $k_1 \neq k_2$, we shall also require - for the analysis in Section 3.3 - that $X_i^{(k_1)}$ is equal to $X_j^{(k_2)}$ in distribution whenever $t_i^{(k_1)} = t_j^{(k_2)}$. To deal with both these problems, we use a special construction to generate the Markov chains $X^{(k)}$.

On the finest time-grid $\pi^{(\kappa)}$, let $(X_i^{(\kappa)})_{0,\leq i\leq 2^\kappa}$ to be a $\pi^{(\kappa)}$-Markov chain that satisfies (3.1.8). This may impose a condition on the size of κ. For example, κ must be sufficiently large so that the Euler approximation $X^{(\kappa)}$ satisfies (3.1.8). We define the $\pi^{(k)}$-Markov chains $X^{(k)}$ for the less refined time-grids $\pi^{(k)}$, with $k<\kappa$, in the following way: for every $i \in \{0,\dots,2^k\}$ there is a $j \in \{0\dots,2^\kappa\}$ with $t_i^{(k)} = t_j^{(\kappa)}$, so we set $X_i^{(k)} = X_j^{(\kappa)}$. This ensures that $X^{(k)}$ satisfies (3.1.8) and that $X_i^{(k)}$ has the same distribution as $X_j^{(\kappa)}$.

On the other hand, assume that we can compute the joint marginals $(X_{t_0^{(k)}},\dots,X_{t_{2^k}^{(k)}})$ of the diffusion X for each time-grid. The canonical examples to think of are $X = W$, or X is a geometric Brownian motion. In this case, we set $X_i^{(k)} = X_{t_i^{(k)}}$, and we avoid the disadvantages of the more general scheme stated above.

3.2 Zero driver-BSDE

In this section, we introduce the multilevel scheme. Sections 3.2.1-3.2.3 contain preparatory theory, and Section 3.3 contains the Monte Carlo scheme. The results of Section 3.2.1 and Section 3.2.3 are valid for (y,z), the 0-driver BSDE in (3.1.1), but not in general for (Y,Z), the full BSDE in (3.1.4). For this reason, we can only apply the multilevel scheme to the 0-driver BSDE.

3.2.1 Theory with general terminal condition

First, we consider BSDEs of the form

$$\tilde{y}_t = \xi - \int_t^T \tilde{z}_s dW_s, \quad 0 \leq t \leq T. \tag{3.2.1}$$

for any $\xi \in \mathbf{L}_2(\mathcal{F}^{(W)})$. It follows from the Martingale Representation Theorem that there exists a unique predictable solution to (3.2.1) in $\mathbf{L}_2([0,T]\times\Omega^{(W)})$. The 0-driver BSDE (y,z) in (3.1.1) is a special case of (3.2.1). Given the time-grid $\pi^{(k)}$ (for any k), the Kunita-Watanabe decomposition [FS04, Theorem 10.17] guarantees us the existence of a unique pair of square integrable, $\pi^{(k)}$-adapted processes, denoted by $(y^{(k)},z^{(k)})$, and a square integrable $(\mathcal{F}_i^{(k)})$-martingale $L^{(k)}$ such that

$$\tilde{y}_i^{(k)} = \xi - \sum_{j=i}^{2^k-1} \tilde{z}_j^{(k)} \Delta W_j^{(k)} - \sum_{j=i}^{2^k-1} \Delta L_j^{(k)} \tag{3.2.2}$$

for every $i \in \{0, \ldots, 2^k\}$, where $\Delta L_i^{(k)} := L_{i+1}^{(k)} - L_i^{(k)}$, $L_0^{(k)} = 0$, and $L^{(k)}$ is "strongly orthogonal" to $W^{(k)}$ in the sense that

$$\mathbb{E}_i^k[\Delta W_i^{(k)} \Delta L_i^{(k)}] = 0 \quad \text{for all } i \in \{0, \ldots, N-1\}.$$

Strong orthogonality essentially implies that $W^{(k)} L^{(k)}$ is an $(\mathcal{F}_i^{(k)})$-martingale. In the following Lemma, we will determine an explicit representation for
$(\tilde{y}^{(k)}, \tilde{z}^{(k)}, L^{(k)})$ in terms of $(\tilde{y}, \tilde{z})$, the solution of the continuous time BSDE (3.2.1), and the conditional expectation $\mathbb{E}_i^k[\cdot]$.

Lemma 3.2.1. *For any time-grid $\pi^{(k)}$, $k \geq 0$, and any time-point $i = 0, \ldots, 2^k$,*

$$\tilde{y}_i^{(k)} = \mathbb{E}_i^k[\xi], \quad \Delta^{(k)} \tilde{z}_i^{(k)} = \mathbb{E}_i^k[\xi \Delta W_i^{(k)}], \tag{3.2.3}$$

$$\Delta L_i^{(k)} = \int_{t_i^{(k)}}^{t_{i+1}^{(k)}} (\tilde{z}_s - \tilde{z}_i^{(k)}) dW_s \tag{3.2.4}$$

$$\Delta^{(k)} \tilde{z}_i^{(k)} = \mathbb{E}_i^k[\int_{t_i^{(k)}}^{t_{i+1}^{(k)}} \tilde{z}_t dt]. \tag{3.2.5}$$

almost surely.

Proof. Taking the conditional expectation $\mathbb{E}_i^k[\cdot]$ (3.2.2) yields that $\tilde{y}_i^{(k)} = \mathbb{E}_i^k[\xi]$ almost surely. Similarly, multiplying by $\Delta W_i^{(k)}$ and taking the conditional expectation $\mathbb{E}_i^k[\cdot]$ in (3.2.2) yields that $\Delta^{(k)} \tilde{z}_i^{(k)} = \mathbb{E}_i^k[\xi \Delta W_i^{(k)}]$, and so we have proven (3.2.3). Multiplying $\tilde{y}_{t_i^{(k)}}$ by $\Delta W_i^{(k)}$, we use (3.2.1) and take conditional expectation $\mathbb{E}_i^k[\cdot]$ to obtain

$$0 = \mathbb{E}_i^k[\xi \Delta W_i^{(k)}] - \mathbb{E}_i^k[\Delta W_i^{(k)} \int_{t_i^{(k)}}^{t_{i+1}^{(k)}} \tilde{z}_s dW_s] = \Delta^{(k)} \tilde{z}_i^{(k)} - \mathbb{E}_i^k[\int_{t_i^{(k)}}^{t_{i+1}^{(k)}} \tilde{z}_t dt]$$

which proves (3.2.5). Finally, we use $\tilde{y}_{t_i^{(k)}} = \xi - \int_{t_i^{(k)}}^{t_{i+1}^{(k)}} \tilde{z}_s dWs$ and $\tilde{y}_{t_i^{(k)}} = \mathbb{E}_i^k[\xi]$ to show that

$$y_i^{(k)} = y_{t_i^{(k)}} = \xi - \sum_{j=i}^{2^k - 1} \tilde{z}_j^{(k)} \Delta W_j^{(k)} - \sum_{j=i}^{2^k-1} \int_{t_j^{(k)}}^{t_{j+1}^{(k)}} (\tilde{z}_s - \tilde{z}_j^{(k)}) dW_s.$$

Therefore, the martingale part $\sum_{j=i}^{2^k-1} \Delta L_j^{(k)}$ has to be given by the stochastic integral above by the uniqueness of the Kunita-Watanabe decomposition. □

Suppose that we have computed the discrete BSDE $(\tilde{y}^{(k-1)}, \tilde{z}^{(k-1)})$ on the time-grid $\pi^{(k-1)}$, and take $i \in \{0, \ldots, 2^k - 3\}$ so that $t_i^{(k)} \in \pi^{(k)}$ and $t_i^{(k)} < t_{2^{k-1}-1}^{(k-1)}$. For every $j > \alpha(i)$, the tower law gives us $\mathbb{E}_i^k[\Delta W_i^{(k)} \tilde{z}_j^{(k-1)} \Delta W_j^{(k-1)}] = \mathbb{E}_i^k[\Delta W_i^{(k)} \mathbb{E}_j^{k-1}[\tilde{z}_j^{(k-1)} \Delta W_j^{(k-1)}]] = 0$. Also, we have that $\mathbb{E}_i^k[\tilde{y}_{\alpha(i)}^{(k-1)} \Delta W_i^{(k)}] = \tilde{y}_{\alpha(i)}^{(k-1)} \mathbb{E}_i^k[\Delta W_i^{(k)}] = 0$.

3.2.2 Differentiability of the 0-driver BSDE

There are deterministic functions $y_t(x) : \mathbb{R}^d \to \mathbb{R}$ and $z_t(x) : \mathbb{R}^d \to \mathbb{R}^q$ such that $y_t(X_t) = y_t$ and $z_t(X_t) = z_t$. $y_t(x)$ is the classical solution of the linear PDE

$$\partial_t u = \frac{1}{2}\sum_{i,j=1}^{d} (\sigma\sigma^\top)_{i,j} \frac{\partial^2}{\partial_{x_i}\partial_{x_j}} u + \sum_{i=1}^{d} b_i \frac{\partial}{\partial_{x_i}} u, \quad u(T,x) = \Phi(x)$$

and $z_t(x) = (\nabla_x y_t(x))^\top \sigma(t,x)$.

Let $n \in \{1,2\}$. It was shown in the proof of [GM10, Lemma 1.1] that $y_t(\cdot)$ is n-times differentiable for all $t \in [0,T)$, and

$$\left.\begin{array}{rcl} \left|\frac{\partial^r}{\partial x_{i_1}\dots\partial x_{i_r}} y_t(x)\right| & \leq & C(T-t)^{(1-r)/2}, \quad \text{for } r = 1,\dots,n \\ \\ \left|\frac{\partial^r}{\partial x_{i_1}\dots\partial x_{i_r}} z_t(x)\right| & \leq & C(T-t)^{-r/2} \quad \text{for } r = 1,\dots,n-1 \end{array}\right\} \text{ for all } t \in [0,T) \tag{3.2.6}$$

for any $i_1,\dots,i_r \in \{1,\dots,d\}$.

The differentiability property (3.2.6) allows us to determine boundedness and Lipschitz continuity properties for the functions $y_t(x)$ and $z_t(x)$, as demonstrated in the following Lemma:

Lemma 3.2.2. *The representation $y_t(x) = \mathbb{E}[\Phi(X_T^{(t,x)})]$ holds.*

Assume that $\Phi(x) \leq C$ for all $x \in \mathbb{R}^d$. Then $|y_t(x)| \leq C$, $|z_t(x)| \leq C$, $|y_t(x_1) - y_t(x_2)| \leq C|x_1 - x_2|$, and $|z_t(x_1) - z_t(x_2)| \leq C|x_1 - x_2|(T-t)^{-1/2}$.

Proof. The representation $y_t(x) = \mathbb{E}[\Phi(X_T^{(t,x)})]$ is the Feymann-Kac formula. The bound $|y_t(x)| \leq C$ follows directly from the bound on $\Phi(x)$. The bound $|z_t(x)| \leq C$ follows from the formula $z_t(x) = (\nabla_x y_t(x))^\top \sigma(t,x)$: (3.2.6) yields $|\nabla_x y_t(x)| \leq C$, and $|\sigma(t,x)| \leq C$ is assumed in $(\mathbf{A}_{\mathbf{b},\sigma})$.

Finally, the Lipschitz continuity of $x \mapsto z_t(x)$ follows from the first order Taylor expansion of $z_t(x)$ and the uniform bound $|\nabla_x z_t(x)| \leq C(T-t)^{-1/2}$. □

The following assumptions allow for derivatives of order greater than 2 for the functions $y_t(\cdot)$ and $z_t(\cdot)$. They will apply only when specifically stated,unlike the assumptions in Section 3.1.2 which apply throughout this chapter. We shall make use of them in in Sections 3.3.3 and 3.4.5.

$(\mathbf{A}_{\mathbf{diff}})$ For some integer n greater than 2, the coefficients of the SDE $b(\cdot)$ and $\sigma(\cdot)$ are n-times continuously differentiable with uniformly bounded derivatives, and the n-th derivatives are Hölder continuous. The constants C in Definition 3.1.1 are allowed additionally to depend on $\|\nabla_x^j b\|_\infty$ and $\|\nabla_x^j \sigma\|_\infty$ for all $j \in \{1,\dots,n\}$. Moreover, the gradient bounds (3.2.6) are valid for derivatives up to order n.

$(\mathbf{A}_{\partial\mathbf{\Phi}})$ The assumption $(\mathbf{A}_{\mathbf{diff}})$ holds and $\Phi(\cdot)$ is n-times continuously differentiable, and the derivatives are uniformly bounded. The constants C in Definition 3.1.1 are allowed additionally to depend on $\|\nabla_x^j b\|_\infty$, $\|\nabla_x^j \sigma\|_\infty$, and $\|\nabla_x^j \Phi\|_\infty$ for all $j \in \{1,\dots,n\}$.

The assumption $(\mathbf{A}_{\partial\mathbf{\Phi}})$ makes implies stronger gradient bounds that (3.2.6). The following Lemma is a standard result, even without uniformly ellipticity $(\mathbf{A}_{\text{u.e.}})$; see, for example, [TT90, Lemma 2].

Lemma 3.2.3. *Under assumption* $(\mathbf{A}_{\partial\mathbf{\Phi}})$*, the functions* $y_t(\cdot)$ *(resp.* $z_t(\cdot)$*) are* n*-times continuously differentiable with uniformly bounded derivatives.*

$$\left.\begin{array}{rcll}\left|\frac{\partial^r}{\partial x_{i_1}\ldots\partial x_{i_r}} y_t(x)\right| & \leq & C & \text{for } r=1,\ldots,n, \\ \\ \left|\frac{\partial^r}{\partial x_{i_1}\ldots\partial x_{i_r}} z_t(x)\right| & \leq & C & \text{for } r=1,\ldots,n-1\end{array}\right\} \quad \text{for all } t\in[0,T).$$

We remark that there are results on the higher derivatives of the form given in $(\mathbf{A_{diff}})$ in [Kus03]. We do not use the results of [Kus03] in this chapter, since the conditions of are rather different to our own and we have not yet understood the explicit dependency of the constants in that paper. On the other hand, it is useful to bear this paper in mind when thinking about a motivation for $(\mathbf{A_{diff}})$.

3.2.3 Markov structure in the discrete setting

We give two discretizations of the 0-driver BSDE (y,z) in (3.1.4).

For any time-grid $\pi^{(k)}$, define $(\tilde{y}^{(k)},\tilde{z}^{(k)})$ to be the solution of the BSDE (3.2.2) with $\xi=\Phi(X_T)$. We combine the result $\Delta^{(k)}\tilde{z}^{(k)}_i=\mathbb{E}^k_i[\int_{t^{(k)}_i}^{t^{(k)}_{i+1}} z_s ds]$ from Lemma 3.2.1 and [GM10, Theorem 1.1] to show that

$$\sum_{k=0}^{2^k-1}\int_{t^{(k)}_i}^{t^{(k)}_{i+1}}\mathbb{E}^W|z_s-\tilde{z}^{(k)}_i|^2 ds\leq C\Delta^{(k)}. \tag{3.2.7}$$

Lemma 3.2.4. *For all* $i\in\{0,\ldots,2^k-1\}$*, there exist measurable, deterministic functions* $\tilde{y}^{(k)}_i:\mathbb{R}^d\to\mathbb{R}$ *and* $\tilde{z}^{(k)}_i:\mathbb{R}^d\mapsto\mathbb{R}^q$ *such that* $\tilde{y}^{(k)}_i=\tilde{y}^{(k)}_i(X_{t^{(k)}_i})$ *and* $\tilde{z}^{(k)}_i=\tilde{z}^{(k)}_i(X_{t^{(k)}_i})$ *almost surely.*

Proof. We have from Section 3.2.1 that $\tilde{y}^{(k)}_i=y_{t^{(k)}_i}$ and from Section 3.2.2 that $y_{t^{(k)}_i}=y_{t^{(k)}_i}(X_{t^{(k)}_i})$ as required. On the other hand, $\tilde{z}^{(k)}_i=\frac{1}{\Delta^{(k)}}\mathbb{E}^k_i[\int_{t^{(k)}_i}^{t^{(k)}_{i+1}} z_s(X_s)ds]$ is a also deterministic, measurable function evaluated at $X_{t^{(k)}_i}$, because [Kal02, Lemma 8.1] implies that the path $\{X_t: t\geq t^{(k)}_i\}$ is independent of $\mathcal{F}^{(k)}_i$ conditionally on $X_{t^{(k)}_i}$. □

On the other hand, let $(y^{(k)},z^{(k)})$ be the discrete BSDE satisfying (3.2.2) with $\xi=\Phi(X^{(k)}_{2^k})$.

Lemma 3.2.5. *There exist deterministic, measurable functions* $y^{(k)}_i:\mathbb{R}^d\to\mathbb{R}$ *and* $z^{(k)}_i:\mathbb{R}^d\to\mathbb{R}^q$ *such that* $y^{(k)}_i=y^{(k)}_i(X^{(k)}_i)$ *and* $z^{(k)}_i=z^{(k)}_i(X^{(k)}_i)$ *almost surely.*

Proof. For every $i\in\{0,\ldots,2^k-1\}$ and $j\in\{i+1,\ldots,2^k\}$, there is a function $f^{(k)}_{i,j}:\Omega^{(W)}\times\mathbb{R}^d\to\mathbb{R}^d$ that is $\sigma(W_t-W_{t^{(k)}_i}:t\geq t^{(k)}_i)\otimes\mathcal{B}(\mathbb{R}^d)$ such that $X^{(k)}_j=f^{(k)}_{i,j}(X^{(k)}_i)$.

Therefore, $y^{(k)}_i=\mathbb{E}^k_i[\Phi(f^{(k)}_{i,N}(X^{(k)}_i))]$ and $\Delta^{(k)}z^{(k)}_i=\mathbb{E}^k_i[\Phi(f^{(k)}_{i,N}(X^{(k)}_i))\Delta W^{(k)}_i]$; it follows then from Lemma 3.6.1 that $y^{(k)}_i(x)=\mathbb{E}[\Phi(f_{i,N}(x))]$ and $\Delta^{(k)}z^{(k)}_i(x)=\mathbb{E}[\Phi(f_{i,N}(x))\Delta W^{(k)}_i]$. □

Lemma 3.2.6. *There is a constant* C *as in Definition 3.1.1 such that* $|y^{(k)}_i(x)|\leq C$, $|z^{(k)}_i(x)|\leq C$, $|y^{(k)}_i(x_1)-y^{(k)}_i(x_2)|\leq C|x_1-x_2|$*, and* $|z^{(k)}_i(x_1)-z^{(k)}_i(x_2)|\leq C|x_1-x_2|\big((T-t^{(k)}_{i+1})\vee\Delta^{(k)}\big)^{-1/2}+C\Delta^{(k)}$ *for all* $x,x_1,x_2\in\mathbb{R}^d$, $k\in\{0,\ldots,\kappa\}$ *and* $i\in\{0,\ldots,2^k-1\}$.

If $(\mathbf{A}_{\partial\mathbf{\Phi}})$ *is in force for* $n=2$*, i.e.* $\Phi(\cdot)$ *is twice continuously differentiable with uniformly bounded derivatives, then there is a constant* C *as in* $(\mathbf{A}_{\partial\mathbf{\Phi}})$ *such that* $|z^{(k)}_i(x_1)-z^{(k)}_i(x_2)|\leq C|x_1-x_2|+C\Delta^{(k)}$ *uniformly in* i *and* k*.*

Proof. We want to use the bounds and Lipschitz properties of the continuous functions $y_{t_i^{(k)}}(x)$ and $z_{t_i^{(k)}}(x)$ to make conclusions about $y_i^{(k)}(x)$ and $z_i^{(k)}(x)$. To do so, define the additional discrete BSDE

$$(y_{t_j^{(k)}}(X_{t_j^{(k)}}^{(t_i^{(k)},x)}), \frac{1}{\Delta^{(k)}}\mathbb{E}_i^k[\int_{t_j^{(k)}}^{t_{j+1}^{(k)}} z_s(X_s^{(t_i^{(k)},x)})ds])_{j\geq i},$$

which - using Lemma 3.2.1 - solves (3.2.2) with $\xi = \Phi(X_T^{(t_i^{(k)},x)})$. Using direct computations with

$$y_{t_i^{(k)}}(x) = \mathbb{E}_i^k[\Phi(X_T^{(t_i^{(k)},x)})] \quad \text{and} \quad \mathbb{E}_i^k[\int_{t_j^{(k)}}^{t_{j+1}^{(k)}} z_s(X_s^{(t_i^{(k)},x)})ds] = \mathbb{E}_i^k[\Phi(X_T^{(t_i^{(k)},x)})\Delta W_j^{(k)}]$$

we obtain

$$\begin{aligned}
|y_{t_i^{(k)}}(x) - y_i^{(k)}(x)|^2 + |\frac{1}{\Delta^{(k)}}\mathbb{E}_i^k[\int_{t_i^{(k)}}^{t_{i+1}^{(k)}} z_s(X_s^{(t_i^{(k)},x)})ds] - z_i^{(k)}(x)|^2\Delta^{(k)} \\
\leq C\mathbb{E}_i^k[|\Phi(X_T^{(t_i^{(k)},x)}) - \Phi(X_{2^k}^{(k,i,x)})|^2] \\
\leq C\mathbb{E}_i^k[|X_T^{(t_i^{(k)},x)} - X_{2^k}^{(k,i,x)}|^2] \leq C\Delta^{(k)} \qquad (3.2.8)
\end{aligned}$$

Now, using Lemma 3.2.2 and the above computation, we obtain

$$\begin{aligned}
|z_i^{(k)}(x)| &\leq C\Delta^{(k)} + \frac{\int_{t_i^{(k)}}^{t_{i+1}^{(k)}} \mathbb{E}_i^k[|z_t(X_t^{(t_i^{(k)},x)})|]dt}{\Delta^{(k)}} \leq C \\
|z_i^{(k)}(x_1) - z_i^{(k)}(x_2)| &\leq C\Delta^{(k)} + \frac{\int_{t_i^{(k)}}^{t_{i+1}^{(k)}} (\mathbb{E}_i^k[|z_t(X_t^{(t_i^{(k)},x_1)}) - z_t(X_t^{(t_i^{(k)},x_2)})|^2])^{1/2}dt}{\Delta^{(k)}} \\
&\leq C\Delta^{(k)} + \frac{\int_{t_i^{(k)}}^{t_{i+1}^{(k)}} \frac{C(\mathbb{E}_i^k[|X_t^{(t_i^{(k)},x_1)} - X_t^{(t_i^{(k)},x_2)}|^2])^{1/2}}{(T-t)^{1/2}}dt}{\Delta^{(k)}} \\
&\leq C\Delta^{(k)} + \frac{C|x_1 - x_2|}{\big((T - t_{i+1}^{(k)}) \vee \Delta^{(k)}\big)^{1/2}}
\end{aligned}$$

Since $y_i^{(k)}(x) = \mathbb{E}_i^k[\Phi(X_{2^k}^{(k,i,x)})]$, the uniform bounds come from the uniform bounds on Φ. Finally, the Lipschitz continuity of $x \mapsto y_i^{(k)}(x)$ follows from Lemma 3.2.2 and (3.2.8):

$$|y_i^{(k)}(x_1) - y_i^{(k)}(x_2)| \leq C\Delta^{(k)} + |y_{t_i^{(k)}}(x_1) - y_{t_i^{(k)}}(x_2)| \leq C\Delta^{(k)} + C|x_1 - x_2|$$

The proof under $(\mathbf{A_{\partial\Phi}})$ is analogous, but we use the uniform gradient bounds of $(\mathbf{A_{\partial\Phi}})$ rather than (3.2.6). □

Finally, the following Lemma will be useful in Section 3.3. Recall the functions $y_t(x)$ and $z_t(x)$ of Section 3.2.2. Later, it will be desirable to switch from $y_i^{(k)}(x)$ (resp. $z_i^{(k)}(x)$) to $y_{t_i^{(k)}}(x)$ (resp. $z_{t_i^{(k)}}(x)$) due to the differentiability of the functions $x \mapsto y_{t_i^{(k)}}(x)$ (resp. $x \mapsto z_{t_i^{(k)}}(x)$).

Lemma 3.2.7. *The approximation error*

$$\max_{0\le i\le 2^k} \mathbb{E}^W[|y_i^{(k)}(X_i^{(k)}) - y_{t_i^{(k)}}(X_{t_i^{(k)}})|^2] + \sum_{i=0}^{2^k-1} \mathbb{E}^W[|z_i^{(k)}(X_i^{(k)}) - z_{t_i^{(k)}}(X_{t_i^{(k)}})|^2]\Delta^{(k)}$$

is bounded above by $C\Delta^{(k)}$ for each $k \in \{0, \dots, \kappa\}$.

(C does not depend on k or on κ.)

Proof. Recall the BSDE $(\tilde{y}^{(k)}, \tilde{z}^{(k)})$ defined at the beginning of this section. Using direct computations, as in (3.2.8), one obtains

$$\max_{0\le i\le 2^k} \mathbb{E}^W[|y_i^{(k)}(X_i^{(k)}) - \tilde{y}_i^{(k)}|^2] + \sum_{i=0}^{2^k-1} \mathbb{E}^W[|z_i^{(k)}(X_i^{(k)}) - \tilde{z}_i^{(k)}]|^2]\Delta^{(k)} \le C\mathbb{E}^W[|\Phi(X_T) - \Phi(X_{2^k}^{(k)})|^2] \le C\Delta^{(k)}.$$

Note that $\tilde{y}_i^{(k)}$ is actually equal to $y_{t_i^{(k)}}$, so the above inequality yields that $\max_{0\le i\le 2^k} \mathbb{E}^W[|y_i^{(k)}(X_i^{(k)}) - y_{t_i^{(k)}}|^2]$ is bounded above by $C\Delta^{(k)}$. On the other hand, $\sum_i \mathbb{E}^W[|z_{t_i^{(k)}}(X_{t_i^{(k)}}) - \tilde{z}_i^{(k)}]|^2]\Delta^{(k)}$ can be bounded above using Jensen's inequality as follows:

$$\begin{aligned}\sum_{i=0}^{2^k-1} \mathbb{E}^W[|z_{t_i^{(k)}}(X_{t_i^{(k)}}) - \tilde{z}_i^{(k)}]|^2]\Delta^{(k)} &\le \sum_{i=0}^{2^k-1} \mathbb{E}^W[|z_{t_i^{(k)}}(X_{t_i^{(k)}}) - \frac{1}{\Delta^{(k)}}\mathbb{E}_i^k[\int_{t_i^{(k)}}^{t_{i+1}^{(k)}} z_t(X_t)dt]|^2]\Delta^{(k)} \\ &\le \frac{1}{\Delta^{(k)}} \int_{t_i^{(k)}}^{t_{i+1}^{(k)}} \mathbb{E}^W[|z_{t_i^{(k)}}(X_{t_i^{(k)}}) - z_t(X_t)|^2]dt\end{aligned}$$

In fact, (3.2.7) is proved by showing exactly that the last inequality above is bounded above by $C\Delta^{(k)}$; see the proof of [GM10, Theorem 1.1] for details. Therefore, it follows that $\sum_i \mathbb{E}^W[|z_{t_i^{(k)}}(X_{t_i^{(k)}}) - \tilde{z}_i^{(k)}]|^2]\Delta^{(k)}$ is bounded above by $C\Delta^{(k)}$. The proof of this lemma is then completed by combining Young's inequality and (3.2.7):

$$\begin{aligned}&\max_{0\le i\le 2^k} \mathbb{E}^W[|y_i^{(k)}(X_i^{(k)}) - y_{t_i^{(k)}}(X_{t_i^{(k)}})|^2] + \sum_{i=0}^{2^k-1} \mathbb{E}^W[|z_i^{(k)}(X_i^{(k)}) - z_{t_i^{(k)}}(X_{t_i^{(k)}})|^2]\Delta^{(k)} \\ &\qquad \le \max_{0\le i\le 2^k} \mathbb{E}^W[|y_i^{(k)}(X_i^{(k)}) - y_{t_i^{(k)}}(X_{t_i^{(k)}})|^2] + 2\sum_{i=0}^{2^k-1} \mathbb{E}^W[|z_{t_i^{(k)}}(X_{t_i^{(k)}}) - \tilde{z}_i^{(k)}]|^2]\Delta^{(k)} \\ &\qquad\qquad + \sum_{i=0}^{2^k-1} \mathbb{E}^W[|z_i^{(k)}(X_i^{(k)}) - \tilde{z}_i^{(k)}]|^2]\Delta^{(k)}\end{aligned}$$

and the last inequality is bounded above by $C\Delta^{(k)}$ from the computations above. □

3.3 Multilevel Monte Carlo scheme

In this section, we construct approximations the functions $y_i^{(k)}(x)$ (resp. $z_i^{(k)}(x)$) of Section 3.2.3 for each level k of the multilevel algorithm and each time-point i of the grid $\pi^{(k)}$. The approximating functions are denoted by

$$y_i^{(k,M)} : \mathbb{R}^d \to \mathbb{R}, \qquad \text{resp. } z_i^{(k,M)} : \mathbb{R}^d \to \mathbb{R}^q. \tag{3.3.1}$$

We then obtain approximations for the discrete BSDE $(y^{(k)}, z^{(k)})$ by evaluating at $X^{(k)}$: $y_i^{(k)} \approx y_i^{(k,M)}(X_i^{(k)})$ and $z_i^{(k)} \approx z_i^{(k,M)}(X_i^{(k)})$.

The construction of the approximating functions follows a recursive procedure up the multilevel structure: having computed the functions $y^{(k,M)}(x)$ and $z^{(k,M)}(x)$ of level k, one uses them as control variates to computes $y^{(k+1,M)}(x)$ and $z^{(k+1,M)}(x)$. The details of the algorithm are given in section 3.3.1. In Section 3.3.2, a comprehensive error analysis gives an upper bound for the strong error

$$\max_{0\le i\le 2^k-1} \mathbb{E}[|y_i^{(k)} - y_i^{(k,M)}(X_i^{(k)})|^2] + \sum_{i=0}^{2^k-1} \mathbb{E}[|z_i^{(k)} - z_i^{(k,M)}(X_i^{(k)})|^2]\Delta^{(k)} \tag{3.3.2}$$

in terms of the numerical parameters - the number of Monte Carlo simulations, the choice of basis functions, the level k - of the algorithm. This error analysis allows us to calibrate the numerical parameters to optimize the algorithm by complexity analysis in Section 3.3.3.

3.3.1 Algorithm

Linear least-squares regression. We denote M i.i.d. samples of a given random variable U by $U^{1:M} = \{U^1, \dots, U^M\}$.

Definition 3.3.1. *Given a set of* basis functions $\{p^{(r)} : \mathbb{R}^d \to \mathbb{R} \ : \ 0 \le r \le K\}$, *$\mathbb{R}^d$-valued random variables $X^{1:M}$ (*observations*), and $\mathbb{R}$-valued random variables $S^{1:M}$ (*response*), recall that the linear least-squares regression computes the coefficients*

$$\left.\begin{array}{rcl} \mathcal{A} &=& \{\alpha \in \mathbb{R}^K \ : \ \mathit{arg}\min_{\beta\in\mathbb{R}^K} \frac{1}{M}\sum_{m=1}^M |\beta \cdot p(X^m) - S^m|^2\} \\ \alpha^* &=& \mathit{arg}\min_{\alpha\in\mathcal{A}} \|\alpha\|_{\mathbb{R}^K} \end{array}\right\} \tag{3.3.3}$$

where $p(x) = (p^{(1)}(x), \dots, p^{(K)}(x))$ and $\|\cdot\|_{\mathbb{R}^K}$ is the usual Euclidean norm.

In implementation, we use a Singular Value Decomposition approach to compute the coefficient α^*; see [GVL96] for a detailed account.

In what follows, we will repeatedly compute linear least-squares regressions. In order to avoid repetition, we will shall say that the coefficient α^* computed by the above procedure solves

$$LS(X^{(1:M)}, S^{(1:M)}, p(x)). \tag{3.3.4}$$

Basis functions. For each $i \in \{0, \dots, 2^\kappa - 1\}$ and $l \in \{0, \dots, q\}$, we are given a set of basis functions $\{p_{u,l,i}^{(\kappa)}(x) \ : \ u = 1, \dots, K_{l,i}^{(\kappa)}\}$. We denote by $p_{l,i}^{(\kappa)}(x)$ the vector $(p_{1,l,i}^{(\kappa)}(x), \dots, p_{K_{l,i}^{(\kappa)},l,i}^{(\kappa)}(x))$. These functions are of the form $p_{u,l,i}^{(\kappa)}(x) = \mathbf{1}_{A_{u,l,i}}(x)$, where $\{A_{u,l,i} \ : \ u = 1, \dots, K_{l,i}^{(\kappa)}\}$ are disjoint sets such that there exists a finite $c_{l,i} \geq 1$ such that

$$\mathbb{P}(X_i^{(\kappa)} \in A_{u,l,i}) \geq \frac{1}{c_{l,i} K_{l,i}^{(\kappa)}} \quad \text{for all } u \in \{1, \ldots, K_{l,i}^{(\kappa)}\}.$$

Basis functions of this form have been proposed by [BG13] in Theorem 2.3 of that paper. This special structure is used to allow bounding of the expectation of the conditional variance rather than the supremum in the error analysis of the regression scheme. Indeed, standard results in regression theory (e.g. [GKKW02, Theorem 11.1]) typically require that $\sup_x Var(S|X = x)$ be bounded, whereas with this choice of basis it is sufficient to work with $\mathbb{E}[Var(S|X)]$ be bounded; this is crucial in the analysis.

In what follows, for given $u, k \in \{0, \ldots, \kappa\}$, $i \in \{0, \ldots, 2^k - 1\}$, $j \in \{0, \ldots, 2^u - 1\}$, and $l \in \{0, \ldots, q\}$, if $t_i^{(k)} = t_j^{(u)}$, then the basis functions $p_{l,i}^{(k)}(x)$ will be the same as $p_{l,j}^{(u)}(x)$.

Simulations. For each $k \in \{0, \ldots, \kappa\}$, let M_k be an integer and generate M_k copies $\{(\Omega^{(k,m)}, \mathcal{F}^{(k,m)}, \mathbb{P}^{(k,m)}) : m = 1, \ldots, M_k\}$ of the probability space $(\Omega^{(W)}, \mathcal{F}^{(W)}, \mathbb{P}^{(W)})$. Define $X^{(k,m)}$ to be the copy of $X^{(k)}$ and $\Delta W^{(k,m)}$ to be the copy of $\Delta W^{(k)}$ in $(\Omega^{(k,m)}, \mathcal{F}^{(k,m)}, \mathbb{P}^{(k,m)})$; we call these objects the *simulations*, because, in practice, one generates these objects using random number generators. Define by $(\Omega, \mathcal{F}, \mathbb{P})$ the product space of $(\Omega^{(W)}, \mathcal{F}^{(W)}, \mathbb{P}^{(W)})$ and $\bigotimes_{k,m}(\Omega^{(k,m)}, \mathcal{F}^{(k,m)}, \mathbb{P}^{(k,m)})$, and $\mathbb{E}$ the associated expectation operator.

The cloud of processes $X^{(k,1:M_k)} = \{X^{(k,m)} : m = 1, \ldots, M_k\}$ and Brownian increments $\Delta W^{(k,1:M_k)}$ are independent copies of $X^{(k)}$ and $\Delta W^{(k)}$, respectively, in $(\Omega, \mathcal{F}, \mathbb{P})$. Moreover, the clouds $\{(X^{(k,1:M_k)}, \Delta W^{(k,1:M_k)}) : k = 0, \ldots, \kappa\}$ are independent in $(\Omega, \mathcal{F}, \mathbb{P})$.

Recall that $\mathcal{F}_i^{(k)} := \sigma(W_t : t \leq t_i^{(k)})$. Define the additional σ-algebras

$$\left.\begin{aligned} \mathcal{F}^{(k,*)} &:= \sigma(X_j^{(\nu,m)}, \Delta W_j^{(\nu,m)} \ : \ \nu < k;\ 0 \leq j \leq 2^\nu\ ;\ 1 \leq m \leq M_\nu), \\ \mathcal{F}_i^{(k,*)} &:= \mathcal{F}^{(k,*)} \vee \sigma(X_i^{(k,m)} \ : \ 1 \leq m \leq M_k), \\ \mathcal{F}^{(*)} &:= \sigma(X_j^{(\nu,m)}, \Delta W_j^{(\nu,m)} \ : \ 0 \leq \nu \leq \kappa\ ;\ 0 \leq j \leq 2^\nu\ ;\ 1 \leq m \leq M_\nu). \end{aligned}\right\} \tag{3.3.5}$$

$\mathcal{F}^{(k,M)}$ is the σ-algebra generated by the simulations for the time-grids that are coarser than $\pi^{(k)}$. $\mathcal{F}^{(*)}$ is the σ-algebra generated by all the simulated data.

Definition 3.3.2. *Let $\mathbb{E}_k^*$ (resp. $\mathbb{P}_k^*$) be the conditional expectation (resp. conditional probability) with respect to $\mathcal{F}^{(k,*)}$, $\mathbb{E}_i^{(k,*)}$ (resp. $\mathbb{P}_i^{(k,*)}$) be the conditional expectation (resp. conditional probability) with respect to the σ-algebra $\mathcal{F}_i^{(k,*)}$, and $\mathbb{E}^*$ (resp. $\mathbb{P}^*$) be the conditional expectation (resp. conditional probability) with respect to the σ-algebra $\mathcal{F}^{(*)}$.*

Finally, let $\mathbb{E}_i^k[\cdot] := \mathbb{E}[\cdot|\mathcal{F}_i^{(k)}]$. The conditional expectation operator $\mathbb{E}_i^k$ was defined in Section 3.1.1 for the probability space $(\Omega^{(W)}, \mathcal{F}^{(W)}, \mathbb{P}^{(W)})$ only, and is now extended to the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Response. Assuming that $(y^{(k-1,M)}, z^{(k-1,M)})$ has been computed, define, for each $i \in$

$\{0,\dots,2^k-1\}$, $l \in \{1,\dots,q\}$, and $m \in \{1,\dots,M_k\}$, the following random variables:

$$\Delta^{(k)}\Xi^{(m)}_{l,k,i} := \Delta W^{(k,m)}_{l,i}\Bigg(\Phi(X^{(k,m)}_{2^k}) - y^{(k-1,M)}_{\alpha(i)}(X^{(k,m)}_{2\alpha(i)}) - \sum_{j=\alpha(i)+1}^{2^{k-1}-1} z^{(k-1,M)}_j(X^{(k,m)}_{2j})(\Delta W^{(k,m)}_{2j} + \Delta W^{(k,m)}_{2j+1})\Bigg) \tag{3.3.6}$$

Regression coefficients and approximate solutions. The multilevel Monte Carlo algorithm is a recursion along the levels $k \in \{0,\dots,\kappa\}$: assuming that $(y^{(k-1,M)}, z^{(k-1,M)})$ has been computed, one computes $(y^{(k,M)}, z^{(k,M)})$, and so on. To initialize the algorithm, compute the approximations of $(y^{(0)}, z^{(0)})$ by standard Monte Carlo:

$$\left.\begin{array}{c} y^{(0,M)}_0(x) = \frac{1}{M_0}\sum_{m=1}^{M_0}\Phi(X^{(0,m)}_1), \qquad y^{(0,M)}_1(x) = \Phi(x), \\ z^{(0,M)}_0(x) = \frac{1}{TM_0}\sum_{m=1}^{M_0}\Phi(X^{(0,m)}_1)\Delta W^{(0,m)}_0. \end{array}\right\}$$

For $k > 0$, $i \in \{0,\dots,2^k-1\}$ and $l \in \{1,\dots,q\}$, the coefficients $\alpha^{(k,M)}_{l,i} \in \mathbb{R}^{K^{(k)}_{l,i}}$ are determined by least-squares regression - Definition 3.3.1 - so that

$$\begin{array}{l} \alpha^{(k,M)}_{0,i} \text{ solves } LS(X^{(k,1:M_k)}_i, \Phi(X^{(k,1:M_k)}_{2^k}), p^{(k)}_{0,i}(x)), \\ \alpha^{(k,M)}_{l,i} \text{ solves } LS(X^{(k,1:M_k)}_i, \Xi^{(1:M_k)}_{l,k,i}, p^{(k)}_{l,i}(x)). \end{array}$$

where the operator $LS(\cdot)$ is given in (3.3.3), and $\Xi^{(1:M_k)}_{l,k,i}$ is given in (3.3.6). We see now, by observing equation (3.3.6), that the approximations $(y^{(k-1,M)}, z^{(k-1,M)})$ serve as control variates for the computation of the regression coefficients.

Note that computing the conditional expectation $\mathbb{E}^k_i[\Psi]$ for any $\Psi \in \mathbf{L}_2(\mathcal{F})$ is equivalent to solving the infinite dimensional minimization problem

$$\arg\min_{\psi\in\mathbf{L}_2(\mathcal{F}^{(k)}_i)}\mathbb{E}[|\Psi-\psi|^2]$$

Computing $\alpha^{(k,M)}_{0,i}$ is the finite-dimensional, empirical analogue of taking the conditional expectation $\mathbb{E}^k_i[\cdot]$ in (3.2.2), and computing $\alpha^{(k,M)}_{l,i}$ is finite-dimensional, empirical analogue of taking the conditional expectation $\mathbb{E}^k_i[\cdot]$ in (3.1.3).

Recall from Lemma 3.2.6 that the functions $y^{(k)}_i(x)$ and $z^{(k)}_i(x)$ are uniformly bounded. Briefly denoting by C_y (resp. C_z) the absolute bound of $y^{(k)}_i(x)$ (resp. $z^{(k)}_i(x)$) given in Lemma 3.2.6, define the truncation operators

$$[x]_y := -C_y \wedge x \vee C_y, \quad [x]_z := -C_z \wedge x \vee C_z, \qquad \forall\, x \in \mathbb{R}.$$

Definition 3.3.3. *For every $k > 0$, $i \in \{0,\dots,2^k-1\}$, and $l \in \{1,\dots,q\}$, define the multilevel Monte Carlo approximations $(y^{(k,M)}_i(x), z^{(k,M)}_i(x))$ of the functions $(y^{(k)}_i(x), z^{(k)}_i(x))$ by*

$$\left.\begin{array}{c} y^{(k,M)}_i(x) := [\alpha^{(k,M)}_{0,i}\cdot p^{(k)}_{0,i}(x)]_y, \quad z^{(k,M)}_{l,i}(x) := [\alpha^{(k,M)}_{l,i}\cdot p^{(k)}_{l,i}(x)]_z, \\ z^{(k,M)}_i(x) := \big(z^{(k,M)}_{1,i}(x),\dots,z^{(k,M)}_{1,i}(x)\big). \end{array}\right\} \tag{3.3.7}$$

The use of truncation prevents excessive error due to overfitting, thereby improving numerical stability. Moreover, having bounded functions allows the use of covering techniques in the error analysis, in particular in Lemma 3.3.5 below. We refer the reader to [GKKW02, Chapter 11] for a general introduction to the use of covering techniques to least-squares regression.

3.3.2 Error analysis

In this section, we determine upper bounds for the error terms

$$\left.\begin{array}{lll}\mathcal{E}(Y,k,i) & := & \mathbb{E}[\frac{1}{M_k}\sum_{m=1}^{M_k}|y_i^{(k)}(X_i^{(k,m)})-y_i^{(k,M)}(X_i^{(k,m)})|^2],\\ \mathcal{E}(Z,k,i) & := & \mathbb{E}[\frac{1}{M_k}\sum_{m=1}^{M_k}|z_i^{(k)}(X_i^{(k,m)})-z_i^{(k,M)}(X_i^{(k,m)})|^2]\end{array}\right\} \tag{3.3.8}$$

in each level of the multilevel scheme terms of the basis functions, the number of simulations and $\sum_{j=0}^{2^{k-1}-1}\mathcal{E}(Z,k-1,j)\Delta^{(k-1)}$. For notational simplicity, we introduce the following *random norms*:

Definition 3.3.4. *Let $f:\Omega\times\mathbb{R}^d\to\mathbb{R}$ be some $\mathcal{F}^{(*)}\otimes\mathcal{B}(\mathbb{R}^d)$-measurable function. For each $k\in\{0,\dots,\kappa\}$ and $i\in\{0,\dots,2^k\}$, define the random operators $\|f\|_{k,i,\infty}^2:=\int|f(x)|^2\mathbb{P}\circ(X_i^{(k)})^{-1}(dx)$ and $\|f\|_{k,i,M}^2:=\frac{1}{M_k}\sum_{m=1}^{M_k}|f(X_i^{(k,m)})|^2$.*

The operators $\|\cdot\|_{k,i,\infty}$ and $\|\cdot\|_{k,i,M}$ are $\mathcal{F}^{(*)}$-measurable. In particular, $\|f\|_{k,i,\infty}=\mathbb{E}^*[f(X_i^{(k)})]$.

Although the error terms (3.3.8) are in terms of the empirical norm $\|\cdot\|_{k,i,M}$, they are related to the usual norm through the inequalities

$$\left.\begin{array}{l}\mathbb{E}[|y_i^{(k)}(X_i^{(k)})-y_i^{(k,M)}(X_i^{(k)})|^2]\le\varepsilon_Y^{(k)}+\mathbb{E}[\|y_i^{(k)}-y_i^{(k,M)}\|_{k,i,M}^2]+C\mathbb{P}(A_{Y,i}^{(k)}),\\ \mathbb{E}[|z_i^{(k)}(X_i^{(k)})-z_i^{(k,M)}(X_i^{(k)})|^2]\le\varepsilon_Z^{(k)}+\mathbb{E}[\|z_i^{(k)}-z_i^{(k,M)}\|_{k,i,M}^2]+C\mathbb{P}(A_{Z,i}^{(k)}).\end{array}\right\} \tag{3.3.9}$$

for free parameters $\varepsilon_Y^{(k)}\ge 0$ and $\varepsilon_Z^{(k)}\ge 0$ - to be chosen in the complexity analysis, Section 3.3.3 - and the events

$$\left.\begin{array}{ll}A_{Y,i}^{(k)} & :=\{\,\varepsilon_Y^{(k)}\le\|y_i^{(k)}-y_i^{(k,M)}\|_{k,i,\infty}^2-\|y_i^{(k)}-y_i^{(k,M)}\|_{k,i,M}^2\},\\ A_{Z,i}^{(k)} & :=\{\exists l\in\{1,\dots,q\}:\varepsilon_Z^{(k)}\le\|z_{l,i}^{(k)}-z_{l,i}^{(k,M)}\|_{k,i,\infty}^2-\|z_{l,i}^{(k)}-z_{l,i}^{(k,M)}\|_{k,i,M}^2\}.\end{array}\right\} \tag{3.3.10}$$

The probabilities $\mathbb{P}(A_{Y,i}^{(k)})$ and $\mathbb{P}(A_{Z,i}^{(k)})$ converge to zero as the number of simulations M_k goes to infinity. The following concentration of measure inequalities give the rate of convergence in terms of the number of basis functions $K_{l,i}^{(k)}$ and the free parameters $\varepsilon_Y^{(k)}$ and $\varepsilon_Z^{(k)}$.

Lemma 3.3.5. *For all $k\in\{0,\dots,\kappa\}$, $i\in\{0,\dots,2^k\}$, there exists a (possibly different) constant C such that, if $\varepsilon_Y^{(k)}+\varepsilon_Z^{(k)}\le C$,*

$$\left.\begin{array}{ll}\mathbb{P}(A_{Y,i}^{(k)}) & \le C\exp\big(-CK_{0,i}^{(k)}\log(\varepsilon_Y^{(k)})-CM_k\varepsilon_Y^{(k)}\big),\\ \mathbb{P}(A_{Z,i}^{(k)}) & \le C\sum_{l=1}^q\exp\big(-CK_{l,i}^{(k)}\log\big(\varepsilon_Z^{(k)}\big)-CM_k\varepsilon_Z^{(k)}\big).\end{array}\right\} \tag{3.3.11}$$

Proof. The bounds are obtained analogously to the bounds on the sets $C_k^{Y,M}$ and $C_k^{Z,M}$ in Theorem 2.4.5 of Chapter 2. □

Remark. The proof of Lemma 3.3.5 requires that the functions $(y^{(k,M)},z^{(k,M)})$ be bounded. Since $\alpha_{l,i}^{(k,M)}\cdot p_{l,i}^{(k)}(x)$ is in general unbounded, the truncation in the definition of the approximating functions (3.3.7) is a technical necessity. This is a usual technique in the regression theory [GKKW02, Chapter 11].

In Proposition 3.3.8 and Theorem 3.3.9, we express the error terms $\mathcal{E}(Y,k,i)$ and $\mathbb{E}(Z,k,i)$ given in (3.3.8) using the terms of the following definition.

Definition 3.3.6. *For each $k \in \{0,\dots,\kappa\}$ and $i \in \{0,\dots,2^k-1\}$, define*

$$\begin{aligned}
T_{1,i}^{(Y,k)} &:= \mathbb{E}[\inf_{\alpha\in\mathbb{R}^{K_{0,i}^{(k)}}} \|\alpha\cdot p_{0,i}^{(k)} - y_i^{(k)}\|_{k,i,M}^2],\\
T_{1,i}^{(Z,k)} &:= \sum_{l=1}^{q} \mathbb{E}[\inf_{\alpha\in\mathbb{R}^{K_{l,i}^{(k)}}} \|\alpha\cdot p_{l,i}^{(k)} - z_{l,i}^{(k)}\|_{k,i,M}^2],\\
T_{2,i}^{(Y,k)} &:= \mathbb{E}[\|(\alpha_{0,i}^{(k,M)} - \mathbb{E}_i^{(k,*)}[\alpha_{0,i}^{(k,M)}])\cdot p_{0,i}^{(k)}\|_{k,i,M}^2],
\end{aligned}$$

As shown in the following proposition, the terms $T_{1,i}^{(Y,k)}$ and $T_{1,i}^{(Z,k)}$ are a measure of *bias* caused by the basis selection, whereas $T_{2,i}^{(Y,k)}$ is a measure of *variance* of the least-squares regression $LS(X_i^{(k,1:M_k)}, \Phi(X_{2^k}^{(k,1:M_k)}), p_{0,i}^{(k)}(x))$.

Proposition 3.3.7. *For every $k \in \{1,\dots,\kappa\}$ and $i \in \{0,\dots,2^k-1\}$,*

$$T_{1,i}^{(Y,k)} \le \inf_{\alpha\in\mathbb{R}^{K_{0,i}^{(k)}}} \mathbb{E}[|y_i^{(k)}(X_i^{(k)}) - \alpha\cdot p_{0,i}^{(k)}(X_i^{(k)})|^2], \tag{3.3.12}$$

$$T_{1,i}^{(Z,k)} \le \sum_{l=1}^{q} \inf_{\alpha\in\mathbb{R}^{K_{l,i}^{(k)}}} \mathbb{E}[|z_{l,i}^{(k)}(X_i^{(k)}) - \alpha\cdot p_{l,i}^{(k)}(X_i^{(k)})|^2], \tag{3.3.13}$$

$$T_{2,i}^{(Y,k)} \le \frac{\mathbb{E}[|\Phi(X_{2^k}^{(k)})|^2]K_{0,i}^{(k)}}{M_k} \le \frac{CK_{0,i}^{(k)}}{M_k} \tag{3.3.14}$$

Proof. The bounds (4.4.20) and (3.3.13) are obtain in exactly the same way as those of the equivalent terms in [LGW06, Proposition 4]. Although also similar to the bound of the equivalent term in [LGW06, Proposition 4], we include the proof of (3.3.14) for the convenience of the reader.

Denote by $P_{0,i}^{(k,M)}$ the $M_k \times K_{0,i}^{(k)}$-dimensional random matrix whose m-th row is $p_{0,i}^{(k)}(X_i^{(k,m)})$. For the purposes of this proof, we assume (without loss of generality) that $(P_{0,i}^{(k,M)})^\top P_{0,i}^{(k,M)}/M_k = Id$. Indeed, in the case that this does not hold, one can scale the basis functions as follows: for each $u \in \{1,\dots,K_{0,i}^{(k)}\}$ let $\rho_{u,k,0,i} = \frac{1}{M_k}\sum_{m=1}^{M_k} p_{u,0,i}^{(k)}(X_i^{(k,m)})^2$, and $\tilde{p}_{u,0,i}^{(k)}(x)$ be the function $p_{u,0,i}^{(k)}(x)\mathbf{1}_{\{\rho_{k,0,i}>0\}}/\sqrt{\rho_{u,k,0,i}}$ with the convention that $0/0 = 0$. Only $\tilde{K}_{0,i}^{(k)} \le K_{0,i}^{(k)}$ of these functions will be non-zero. Set $\tilde{p}_{0,i}^{(k)}(x)$ to be the $\tilde{K}_{0,i}^{(k)}$-dimensional vector of functions whose j-th component is $\tilde{p}_{u_j,0,i}^{(k)}(x)$, where $\tilde{p}_{u_j,0,i}^{(k)}(x)$ belongs to the set of scaled basis functions that are non-zero. Let $\tilde{\alpha}_{0,i}^{(k,M)}$ solve $LS(X_i^{(k,1:M_k)}, \Phi(X_{2^k}^{(k,1:M_k)}), \tilde{p}_{0,i}^{(k)}(x))$, and observe that $\tilde{\alpha}_{0,i}^{(k,M)}\cdot\tilde{p}_{0,i}^{(k)}(X_i^{(k,m)}) = \alpha_{0,i}^{(k,M)}\cdot p_{0,i}^{(k)}(X_i^{(k,m)})$ for all $m \in \{1,\dots,M_k\}$ because the components of $\tilde{p}_{0,i}^{(k)}(x)$ are just the scaled components of $p_{0,i}^{(k)}(x)$. Moreover, denoting by $\tilde{P}_{0,i}^{(k,M)}$ the $M_k \times K_{0,i}^{(k)}$-dimensional random matrix whose m-th row is $\tilde{p}_{0,i}^{(k)}(X_i^{(k,m)})$, $(\tilde{P}_{0,i}^{(k,M)})^\top\tilde{P}_{0,i}^{(k,M)}/M_k = Id$ follows from the fact that the scaled basis functions are indicators on disjoint sets. Therefore, one could continue the proof below using $\tilde{p}_{0,i}^{(k)}(\cdot)$ in the place of $p_{0,i}^{(k)}(\cdot)$, $\tilde{\alpha}_{0,i}^{(k,M)}$ in the place of $\alpha_{0,i}^{(k,M)}$, $\tilde{K}_{l,i}^{(k)}$ in the place of $K_{0,i}^{(k)}$, and $\tilde{P}_{0,i}^{(k,M)}$ in the place of $P_{0,i}^{(k,M)}$.

Denoting by V the M_k-dimensional random vector whose m-th coordinate is $\Phi(X_{2^k}^{(k,m)})$, the *Normal Equations* [GVL96, Section 5.3.1] yield that

$$p_{0,i}^{(k)}(X_i^{(k,m)}) \cdot (\alpha_{0,i}^{(k,M)} - \mathbb{E}_i^{(k,*)}[\alpha_{0,i}^{(k,M)}]) = (p_{0,i}^{(k)}(X_i^{(k,m)}))^\top \frac{(P_{0,i}^{(k,M)})^\top}{M_k}(V - \mathbb{E}_i^{(k,*)}V),$$

whence, from the $\mathcal{F}_i^{(k,*)}$-measurability of $p_{0,i}^{(k)}(X_i^{(k,m)})$ and $P_{0,i}^{(k,M)}$, it follows that

$$\begin{aligned}&\mathbb{E}_i^{(k,*)}[|p_{0,i}^{(k)}(X_i^{(k,m)}) \cdot (\alpha_{0,i}^{(k,M)} - \mathbb{E}_i^{(k,*)}[\alpha_{0,i}^{(k,M)}])|^2] \\ &= (p_{0,i}^{(k)}(X_i^{(k,m)}))^\top \frac{(P_{0,i}^{(k,M)})^\top}{M_k}\mathbb{E}_i^{(k,*)}[(V - \mathbb{E}_i^{(k,*)}V)(V - \mathbb{E}_i^{(k,*)}V)^\top]\frac{P_{0,i}^{(k,M)}}{M_k}p_{0,i}^{(k)}(X_i^{(k,m)})\end{aligned}$$

Due to the independence of the simulations, the off-diagonal terms of the matrix $\mathbb{E}_i^{(k,*)}[(V - \mathbb{E}_i^{(k,*)}V)(V - \mathbb{E}_i^{(k,*)}V)^\top]$ are zero. On the other hand, the diagonal terms are bounded by C. Therefore, using the Tower Law of conditional expectation, $T_{2,i}^{(Y,k)}$ is bounded by $C\mathbb{E}[\|\frac{P_{0,i}^{(k,M)}}{M_k}p_{0,i}^{(k)}\|_{k,i,M}^2]$. Since $(P_{0,i}^{(k,M)})^\top P_{0,i}^{(M)}/M_k = Id$, the proof is completed by the equality

$$\|\frac{P_{0,i}^{(k,M)}}{M_k}p_{0,i}^{(k)}\|_{k,i,M}^2 = \frac{\operatorname{trace}\big((P_{0,i}^{(k,M)})^\top P_{0,i}^{(k,M)}\big)}{M_k^2} = \frac{K_{0,i}^{(k)}}{M_k}$$

□

Proposition 3.3.8. *For every $k \in \{0, \dots, \kappa\}$ and $i \in \{0, \dots, 2^k - 1\}$,*

$$\mathcal{E}(Y, k, i) = \mathbb{E}[\|y_i^{(k)} - y_i^{(k,M)}\|_{k,i,M}^2] \le T_{1,i}^{(Y,k)} + T_{2,i}^{(Y,k)}. \tag{3.3.15}$$

Proof. Using that $\mathbb{E}_i^{(k,*)}[\Phi(X_{2^k}^{(k,m)}] = y_i^{(k)}(X_i^{(k,m)})$. From Proposition 3.6.4(iii), it follows that $\mathbb{E}_i^{(k,*)}[\alpha_{0,i}^{(k,M)}]$ solves $LS(X_i^{(k,1:M_k)}, y_i^{(k)}(X_i^{(k,1:M_k)}), p_{0,i}^{(k)}(x))$ - see Definition 3.3.1 and (3.3.3) for the operator $LS(\cdot)$. Using Pythagorus' Theorem and Proposition 3.6.4(ii), it follows that

$$\|y_i^{(k)} - y_i^{(k,M)}\|_{k,i,M}^2 \le \|y_i^{(k)} - \mathbb{E}_k^*[\alpha_{0,i}^{(k,M)}] \cdot p_i^{(k)}\|_{k,i,M}^2 + \|(\mathbb{E}_k^*[\alpha_{0,i}^{(k,M)}] - \alpha_{0,i}^{(k,M)}) \cdot p_i^{(k)}\|_{k,i,M}^2$$

Taking expectations yields the result. □

Theorem 3.3.9. *Suppose that $\varepsilon_Y^{(k)} + \varepsilon_Z^{(k)} \le C$ for all k. There exists a (possibly different) constant C such that for every $k \in \{1, \dots, \kappa\}$, $i \in \{0, \dots, 2^k - 1\}$, and $R \ge 1$, the error term $\mathcal{E}(Z, k, i) :=$*

$\mathbb{E}[\|z_i^{(k)} - z_i^{(k,M)}\|_{k,i,M}^2]$ *is bounded by*

$$\frac{C\sum_{l=1}^q K_{l,i}^{(k)}}{\Delta^{(k)}M_k}\Big(\varepsilon_Y^{(k-1)} + T_{1,\alpha(i)}^{(Y,k-1)} + T_{2,\alpha(i)}^{(Y,k-1)} + R\sum_{j=\alpha(i)+1}^{2^{k-1}-1}\mathcal{E}(Z,k-1,j)\Delta^{(k-1)}\Big)$$
$$+\frac{C(R\Delta^{(k)} + e^{-R/2} + R\varepsilon_Z^{(k-1)})\sum_{l=1}^q K_{l,i}^{(k)}}{\Delta^{(k)}M_k} + CT_{1,i}^{(Z,k)}$$
$$+\frac{C\sum_{l=1}^q K_{l,i}^{(k)}}{\Delta^{(k)}M_k}\exp\big(-CK_{0,\alpha(i)}^{(k-1)}\ln(\varepsilon_Y^{(k-1)}) - CM_{k-1}\varepsilon_Y^{(k-1)}\big)$$
$$+\frac{CR\sum_{l=1}^q K_{l,i}^{(k)}}{\Delta^{(k)}M_k}\sum_{j=\alpha(i)+1}^{2^{k-1}-1}\sum_{l=1}^q\exp\big(-CK_{l,j}^{(k-1)}\ln\big(\varepsilon_Z^{(k-1)}\big) - CM_{k-1}\varepsilon_Z^{(k-1)}\big)\Delta^{(k-1)} \quad (3.3.16)$$

Remark. In the first line of (3.3.16) above, we see that $\mathcal{E}(Z,k,i)$ depends on the error terms $\mathcal{E}(Z,k-1,\cdot)$ of the previous level: by reducing the error of level $k-1$, we reduce the error in level k. We make use of this to optimize algorithm efficiency in Section 3.3.3.

Proof. Fix $l\in\{1,\dots,q\}$ and recall the σ-algebras defined in (3.3.5). Consider $\mathbb{E}_i^{(k,*)}[\Xi_{l,k,i}^{(m)}]$, where $\Xi_{l,k,i}^{(m)}$ is defined in (3.3.6). The random functions $\big(y^{(k-1,M)}(x), z^{(k-1,M)}(x)\big)$ are $\mathcal{F}^{(k,*)}$-measurable, whence the random variables $\big(y_{\alpha(j)}^{(k-1,M)}(X_j^{(k,m)}), z_{\alpha(j)}^{(k-1,M)}(X_j^{(k,m)})\big)$ are $\mathcal{F}_j^{(k,*)}$-measurable for every $j\in\{i,\dots,2^k-1\}$ and $m\in\{1,\dots,M_k\}$. This implies that $\mathbb{E}_i^{(k,*)}[\Delta W_{l,i}^{(k,m)} y_{\alpha(i)}^{(k-1,M)}(X_{2\alpha(i)}^{(k,m)})] = 0$. Moreover, using the Tower Law of conditional expectations, $\mathbb{E}_i^{(k,*)}[\Delta W_{l,i}^{(k,m)} z_j^{(k-1,M)}(X_{2j}^{(k,m)})\Delta W_r^{(k,m)}] = 0$ for every $j\in\{\alpha(i)+1,\dots,2^{k-1}-1\}$ and $r\in\{2j,2j+1\}$. Therefore, $\mathbb{E}_i^{(k,*)}[\Xi_{l,k,i}^{(m)}]$ is equal to $\mathbb{E}_i^{(k,*)}[\frac{\Phi(X_{2^k}^{(k,m)})\Delta W_{l,i}^{(k,m)}}{\Delta^{(k)}}]$, which is the l-th component of $z_i^{(k)}(X_i^{(k,m)})$. From Proposition 3.6.4(iii), it follows that $\mathbb{E}_k^*[\alpha_{l,i}^{(k,M)}]$ solves $LS(X_i^{(k,1:M_k)}, z_i^{(k)}(X_i^{(k,1:M_k)}), p_{l,i}^{(k)}(x))$. Pythagoras' Theorem, together with Proposition 3.6.4(ii), then yields

$$\|z_{l,i}^{(k)} - z_{l,i}^{(k,M)}\|_{k,i,M} \le \|z_{l,i}^{(k)} - \mathbb{E}_i^{(k,*)}[\alpha_{l,i}^{(k,M)}]\cdot p_{l,i}^{(k)}\|_{k,i,M} + \|(\mathbb{E}_i^{(k,*)}[\alpha_{l,i}^{(k,M)}] - \alpha_{l,i}^{(k,M)})\cdot p_{l,i}^{(k)}\|_{k,i,M} \quad (3.3.17)$$

Next, we consider $\|(\mathbb{E}_i^{(k,*)}[\alpha_{l,i}^{(k,M)}] - \alpha_{l,i}^{(k,M)})\cdot p_{l,i}^{(k)}\|_{k,i,M}$. To do so, we employ the arguments of the proof of Case (b) in [BG13, Theorem 2.3], which we detail here for the benefit of the reader; observe that the basis functions Φ_k used by the authors in that proof correspond to our own. We start by making a simple rescaling of the basis functions as follows: for each $u\in\{1,\dots,K_{l,i}^{(k)}\}$ let $\rho_{u,k,l,i} = \frac{1}{M_k}\sum_{m=1}^{M_k} p_{u,l,i}^{(k)}(X_i^{(k,m)})^2$, and $\tilde{p}_{u,l,i}^{(k)}(x)$ be the function $p_{u,l,i}^{(k)}(x)\mathbf{1}_{\{\rho_{k,l,i}>0\}}/\sqrt{\rho_{u,k,l,i}}$ with the convention that $0/0=0$. Only $\tilde{K}_{l,i}^{(k)} \le K_{l,i}^{(k)}$ of these functions will be non-zero. Set $\tilde{p}_{l,i}^{(k)}(x)$ to be the $\tilde{K}_{l,i}^{(k)}$-dimensional vector of functions whose j-th component is $\tilde{p}_{u_j,l,i}^{(k)}(x)$, where $\tilde{p}_{u_j,l,i}^{(k)}(x)$ belongs to the set of scaled basis functions that are non-zero. Let $\tilde{\alpha}_{l,i}^{(k,M)}$ solve $LS(X_i^{(k,1:M_k)}, \Xi_{l,k,i}^{(1:M_k)}, \tilde{p}_{l,i}^{(k)}(x))$, and observe that $\tilde{\alpha}_{l,i}^{(k,M)}\cdot\tilde{p}_{l,i}^{(k)}(X_i^{(k,m)}) = \alpha_{l,i}^{(k,M)}\cdot p_{l,i}^{(k)}(X_i^{(k,m)})$ for all $m\in\{1,\dots,M_k\}$ because the components of $\tilde{p}_{l,i}^{(k)}(x)$ are just the scaled components of $p_{l,i}^{(k)}(x)$. Moreover, denoting by $P_{l,i}^{(k,M)}$ the $M_k\times\tilde{K}_{l,i}^{(k)}$-dimensional random matrix whose m-th row is $\tilde{p}_{l,i}^{(k)}(X_i^{(k,m)})$, $(P_{l,i}^{(k,M)})^\top P_{l,i}^{(k,M)}/M_k = Id$ follows from the fact that the scaled basis functions are (scaled) indicator functions on disjoint sets.

As in the proof of Proposition 3.3.7, for every $m\in\{1,\dots,M_k\}$, $(\mathbb{E}_k^*[\tilde{\alpha}_{l,i}^{(k,M)}] - \tilde{\alpha}_{l,i}^{(k,M)})\,\cdot$

$\tilde{p}_{l,i}^{(k)}(X_i^{(k,m)})$ is equal to $(\tilde{p}_{l,i}^{(k)}(X_i^{(k,m)}))^\top \frac{(P_{l,i}^{(k,M)})^\top}{M_k} V$ where V is the M_k-dimensional random vector whose m-th element is $\Xi_{l,k,i}^{(m)} - \mathbb{E}_i^{(k,*)}[\Xi_{l,k,i}^{(m)}]$. Therefore, $\|(\mathbb{E}_k^*[\tilde{\alpha}_{l,i}^{(k,M)}] - \tilde{\alpha}_{l,i}^{(k,M)}) \cdot \tilde{p}_{l,i}^{(k)}\|_{k,i,M}^2$ is equal to

$$V^\top P_{l,i}^{(k,M)} \frac{\sum_{m=1}^{M_k} \left\{ \tilde{p}_{l,i}^{(k)}(X_i^{(k,m)})(\tilde{p}_{l,i}^{(k)}(X_i^{(k,m)}))^\top \right\}}{M_k^3} (P_{l,i}^{(k,M)})^\top V$$

The matrix $\sum_{m=1}^{M_k} \left\{ \tilde{p}_{l,i}^{(k)}(X_i^{(k,m)})(\tilde{p}_{l,i}^{(k)}(X_i^{(k,m)}))^\top \right\} / M_k$ is equal to $(P_{l,i}^{(k,M)})^\top P_{l,i}^{(k,M)} / M_k$ which is the identity. The (m_1, m_2)-th component of the matrix $P_{l,i}^{(k,M)}(P_{l,i}^{(k,M)})^\top$ is equal to $p_{l,i}^{(k)}(X_i^{(k,m_1)}) \cdot p_{l,i}^{(k)}(X_i^{(k,m_2)})$. Taking the conditional expectation, $\mathbb{E}_i^{(k,*)}[V^\top P_{l,i}^{(k,M)}(P_{l,i}^{(k,M)})^\top V]$ is equal to

$$\sum_{m_1,m_2=1}^{M_k} \tilde{p}_{l,i}^{(k)}(X_i^{(k,m_1)}) \cdot \tilde{p}_{l,i}^{(k)}(X_i^{(k,m_2)}) \mathbb{E}_i^{(k,*)}[V_{m_1} V_{m_2}].$$

Due to the independence of the samples, $\mathbb{E}_i^{(k,*)}[V_{m_1} V_{m_2}]$ is nonzero only for $m_1 = m_2$. This implies that $\mathbb{E}_i^{(k,*)}[\|(\mathbb{E}_k^*[\tilde{\alpha}_{l,i}^{(k,M)}] - \tilde{\alpha}_{l,i}^{(k,M)}) \cdot \tilde{p}_{l,i}^{(k)}\|_{k,i,M}^2]$ is bounded by

$$\frac{1}{M_k^2} \sum_{m=1}^{M_k} \mathbb{E}_i^{(k,*)}\big[|V_m|^2\big] |\tilde{p}_{l,i}^{(k)}(X_i^{(k,m)})|^2.$$

Recall that $p_{u,l,i}^{(k)}(x)$ is an indicator on a set that we denote $A_{u,l,i}^{(k)}$ in what follows. By the definition of $\tilde{p}_{l,i}^{(k)}$, $|\tilde{p}_{l,i}^{(k)}(X_i^{(k,m)})|^2$ is equal to $\sum_{j=1}^{\tilde{K}_{l,i}^{(k)}} \mathbf{1}_{A_{u_j,l,i}^{(k)}}(X_i^{(k,m)}) / \rho_{u_j,k,l,i}$, and, by symmetry, it follows that

$$\begin{aligned}
\frac{1}{M_k^2} \sum_{m=1}^{M_k} \mathbb{E}\Big[\mathbb{E}_i^{(k,*)}\big[|V_m|^2\big] |\tilde{p}_{l,i}^{(k)}(X_i^{(k,m)})|^2\Big] &= \frac{1}{M_k} \mathbb{E}\Big[\mathbb{E}\big[|V_1|^2 | X_i^{(k,1)}\big] \sum_{j=1}^{\tilde{K}_{l,i}^{(k)}} \mathbf{1}_{A_{u_j,l,i}^{(k)}}(X_i^{(k,m)}) / \rho_{u_j,k,l,i}\Big] \\
&= \sum_{j=1}^{\tilde{K}_{l,i}^{(k)}} \mathbb{E}\Big[\mathbb{E}\big[|V_1|^2 | X_i^{(k,1)}\big] \mathbf{1}_{A_{u_j,l,i}^{(k)}}(X_i^{(k,1)}) \mathbb{E}[\frac{1}{1 + \sum_{m=2}^{M_k} \mathbf{1}_{A_{u_j,l,i}^{(k)}}(X_i^{(k,m)})}]\Big]
\end{aligned}$$

Since $\sum_{m=2}^{M_k} \mathbf{1}_{A_{u_j,l,i}^{(k)}}(X_i^{(k,m)})$ is binomially distributed with parameters $(M_k - 1, \mathbb{P}(X_i^{(k)} \in A_{u_j,l,i}^{(k)}))$, it follows from [GKKW02, Lemma 4.1] that

$$\mathbb{E}[\frac{1}{1 + \sum_{m=2}^{M_k} \mathbf{1}_{A_{u_j,l,i}^{(k)}}(X_i^{(k,m)})}] \le \frac{1}{M_k \mathbb{P}(X_i^{(k)} \in A_{u_j,l,i}^{(k)})} \le \frac{c_{l,i} K_{l,i}^{(k)}}{M_k}$$

Now, taking the expectation $\mathbb{E}$, and using the Tower Law for conditional expectations to show that $\mathbb{E}\big[\mathbb{E}_i^{(k,*)}[\cdots]\big] = \mathbb{E}\big[\mathbb{E}_k^*[\cdots]\big]$, it follows that

$$\mathbb{E}\Big[\|(\mathbb{E}_k^*[\alpha_{l,i}^{(k,M)}] - \alpha_{l,i}^{(k,M)}) \cdot p_{l,i}^{(k)}\|_{k,i,M}^2\Big] \le \frac{C K_{l,i}^{(k)} \mathbb{E}[\mathcal{R}_{l,i}^{(M)}]}{(\Delta^{(k)})^2 M_k} \tag{3.3.18}$$

for the random variable $\mathcal{R}_{l,i}^{(M)}$ defined by

$$\mathbb{E}_i^*[\Big\{\Big(\Phi(X_i^{(k)}) - y_{\alpha(i)}^{(k-1,M)}(X_{2\alpha(i)}^{(k)}) - \sum_{j=\alpha(i)+1}^{2^{k-1}-1} z_j^{(k-1,M)}(X_{2j}^{(k)}) \cdot (\Delta W_{2j}^{(k)} + \Delta W_{2j+1}^{(k)})\Big)\Delta W_{l,i}^{(k)}\Big\}^2]. \tag{3.3.19}$$

It follows from the triangle inequality that $(\mathbb{E}[\mathcal{R}_{l,i}^{(M)}])^{\frac{1}{2}} \leq (\mathbb{E}[\mathcal{R}_{l,i}^{(1,M)}])^{\frac{1}{2}} + (\mathcal{R}_{l,i}^{(2,M)})^{\frac{1}{2}}$, where

$$\begin{aligned}
\mathcal{R}_{l,i}^{(1,M)} &:= C\mathbb{E}_k^*[\Big\{\Big(y_{\alpha(i)}^{(k-1)}(X_{2\alpha(i)}^{(k)}) - y_{\alpha(i)}^{(k-1,M)}(X_{2\alpha(i)}^{(k)})\Big)\Delta W_{l,i}^{(k)}\Big\}^2] \\
&\quad + C\mathbb{E}_k^*[\Big\{\Delta W_{l,i}^{(k)} \sum_{j=\alpha(i)+1}^{2^{k-1}-1} (z_j^{(k-1,M)}(X_{2j}^{(k)}) - z_j^{(k-1)}(X_{2j}^{(k)})) \cdot (\Delta W_{2j}^{(k)} + \Delta W_{2j+1}^{(k)})\Big\}^2], \\
\mathcal{R}_{l,i}^{(2,M)} &:= C\mathbb{E}[\Big\{\Big(\Phi(X_i^{(k)}) - y_{\alpha(i)}^{(k-1)}(X_{2\alpha(i)}^{(k)}) \\
&\qquad + \sum_{j=\alpha(i)+1}^{2^{k-1}-1} z_j^{(k-1)}(X_{2j}^{(k)}) \cdot (\Delta W_{2j}^{(k)} + \Delta W_{2j+1}^{(k)})\Big)\Delta W_{l,i}^{(k)}\Big\}^2].
\end{aligned}$$

Minkowski's inequality and the independence of the increments and components of the Brownian Motion yield

$$\begin{aligned}
(\mathcal{R}_{l,i}^{(1,M)})^{\frac{1}{2}} &\leq \|y_{\alpha(i)}^{(k-1)} - y_{\alpha(i)}^{(k-1,M)}\|_{k-1,\alpha(i),\infty}(\Delta^{(k)})^{\frac{1}{2}} \\
&\quad + \Big(\sum_{j=\alpha(i)+1}^{2^{k-1}-1}\sum_{r=1}^{q} \mathbb{E}_k^*[(\Delta W_{l,i}^{(k)})^2\big(z_{r,j}^{(k-1)}(X_{2j}^{(k)}) - z_{r,j}^{(k-1,M)}(X_{2j}^{(k)})\big)^2]\Delta^{(k-1)}\Big)^{\frac{1}{2}}
\end{aligned} \tag{3.3.20}$$

Using the bounds of Lemma 3.2.6 and the event $A_{Y,\alpha(i)}^{(k-1)}$ given in (3.3.10), the terms in $\|y_{\alpha(i)}^{(k-1)} - y_{\alpha(i)}^{(k-1,M)}\|_{k-1,\alpha(i),\infty}^2$ are bounded by $\varepsilon_Y^{(k-1)} + \|y_{\alpha(i)}^{(k-1)} - y_{\alpha(i)}^{(k-1,M)}\|_{k-1,\alpha(i),M}^2 + C\mathbf{1}_{(A_{Y,\alpha(i)}^{(k-1)})}$; taking the expectation $\mathbb{E}$ and substituting the bounds (3.3.15) into (3.3.21) for $\mathbb{E}[\|y_{\alpha(i)}^{(k-1)} - y_{\alpha(i)}^{(k-1,M)}\|_{k-1,\alpha(i),M}^2]$ from Proposition 3.3.8, one obtains that

$$\mathbb{E}[\|y_{\alpha(i)}^{(k-1)} - y_{\alpha(i)}^{(k-1,M)}\|_{k-1,\alpha(i),\infty}^2] \leq \varepsilon_Y^{(k-1)} + T_{1,\alpha(i)}^{(Y,k-1)} + T_{2,\alpha(i)}^{(Y,k-1)} + C\mathbf{1}_{(A_{Y,\alpha(i)}^{(k-1)})} \tag{3.3.21}$$

To treat the terms in z in (3.3.20), one decomposes the Brownian increment $\Delta W_i^{(k)}$ into a bounded and unbounded part. Introducing a free parameter $R > 0$ - to be chosen in the complexity analysis, Section 3.3.3 - a Gaussian random variable $\mathcal{N}$ with mean 0 and variance 1, and the truncation operator $\phi_R(x) = -\sqrt{R} \vee x \wedge \sqrt{R}$, one concludes from Mill's inequality that

$$\mathbb{E}[\big(\mathcal{N} - \phi_R(\mathcal{N})\big)^2] = 2\big(\mathbb{P}(\mathcal{N} > \sqrt{R})(R+1) - \frac{\sqrt{R}e^{-R/2}}{\sqrt{2\pi}}\big) \leq 2\mathbb{P}(\mathcal{N} > \sqrt{R})(R+1-R) \leq 2e^{-R/2} \tag{3.3.22}$$

Using the bounds of Lemma 3.2.6 on $|z_{r,j}^{(k-1)}(x)|$,

$$\begin{aligned}
&\mathbb{E}_k^*[(\Delta W_i^{(k)})^2\big(z_{r,j}^{(k-1)}(X_{2j}^{(k)}) - z_{r,j}^{(k-1,M)}(X_{2j}^{(k)})\big)^2] \\
&\quad \le C\mathbb{E}[\big(\Delta W_i^{(k)} - \phi_{R\Delta^{(k)}}(\Delta W_i^{(k)})\big)^2] + CR\Delta^{(k)}\mathbb{E}_k^*[\big(z_{r,j}^{(k-1)}(X_{2j}^{(k)}) - z_{r,j}^{(k-1,M)}(X_{2j}^{(k)})\big)^2] \\
&\quad = C\Delta^{(k)}\mathbb{E}[\big(\mathcal{N} - \phi_R(\mathcal{N})\big)^2] + CR\Delta^{(k)}\|z_{r,j}^{(k-1)} - z_{r,j}^{(k-1,M)}\|_{k-1,j,\infty}^2 \\
&\quad \le C\Delta^{(k)}e^{-R/2} + CR\Delta^{(k)}(\varepsilon_Z^{(k-1)} + \|z_{r,j}^{(k-1)} - z_{r,j}^{(k-1,M)}\|_{k-1,j,M}^2 + C\mathbf{1}_{A_{Z,j}^{(k-1)}})
\end{aligned} \tag{3.3.23}$$

Taking the expectation $\mathbb{E}$ in 3.3.20 and substituting (3.3.21) and (3.3.23), one obtains the bound

$$\begin{aligned}
(\Delta^{(k)})^{-1}\mathbb{E}[\mathcal{R}_{l,i}^{(1,M)}] \le{}& C\big(\varepsilon_Y^{(k-1)} + T_{1,\alpha(i)}^{(Y,k-1)} + T_{2,\alpha(i)}^{(Y,k-1)} + \mathbb{P}(A_{Y,\alpha(i)}^{(k-1)}) + e^{-R/2} + R\varepsilon_Z^{(k-1)}\big) \\
&+ CR\sum_{j=\alpha(i)+1}^{2^{k-1}-1}\Big(\sum_{r=1}^{q}\mathbb{E}[\|z_{r,j}^{(k-1)} - z_{r,j}^{(k-1,M)}\|_{k-1,j,M}^2] + \mathbb{P}(A_{Z,j}^{(k-1)})\Big)\Delta^{(k-1)}
\end{aligned} \tag{3.3.24}$$

Let $(\tilde{y},\tilde{z})$ be the solution of the solution of the the BSDE (3.2.1) with terminal condition $\xi = \Phi(X_{2^k}^{(k)})$. Recalling the results of Section 3.2.3, $(y^{(k-1)}, z^{(k-1)})$ solves a discrete BSDE of the form (3.2.2) on $\pi^{(k-1)}$ with terminal condition $\xi = \Phi(X_{2^k}^{(k)})$. Rearranging (3.2.2) yields

$$\Phi(X_{2^k}^{(k)}) - y_{\alpha(i)}^{(k-1)} - \sum_{j=\alpha(i)+1}^{2^{k-1}-1} z_j^{(k-1)}\Delta W_j^{(k-1)} = z_{\alpha(i)}^{(k-1)}\Delta W_{\alpha(i)}^{(k-1)} + \sum_{j=\alpha(i)}^{2^{k-1}-1}\Delta L_j^{(k-1)}. \tag{3.3.25}$$

Moreover, $\Delta L_j^{(k-1)} = \int_{t_j^{(k-1)}}^{t_{j+1}^{(k-1)}}(\tilde{z}_t - z_j^{(k-1)})dW_t$ due to (3.2.4) in Lemma 3.2.1. Substituting (3.3.25) into the definition of $\mathcal{R}_{l,i}^{(2,M)}$, one obtains

$$\begin{aligned}
\mathcal{R}_{l,i}^{(2,M)} &= \mathbb{E}[\Big\{\Big(\Phi(X_{2^{k-1}}^{(k-1)}) - y_{\alpha(i)}^{(k-1)}(X_{2\alpha(i)}^{(k)}) - \sum_{j=\alpha(i)+1}^{2^{k-1}-1} z_j^{(k-1)}(X_{2j}^{(k)})(\Delta W_{2j}^{(k)} + \Delta W_{2j+1}^{(k)})\Big)\Delta W_{l,i}^{(k)}\Big\}^2] \\
&= \mathbb{E}[\Big\{\Big(\sum_{j=\alpha(i)}^{2^{k-1}-1}\Delta L_j^{(k-1)} + z_{\alpha(i)}^{(k-1)}(X_{2\alpha(i)}^{(k)})\cdot(\Delta W_{2\alpha(i)}^{(k)} + \Delta W_{2\alpha(i)+1}^{(k)})\Big)\Delta W_{l,i}^{(k)}\Big\}^2] \\
&= \mathbb{E}[\big(\Delta W_{l,i}^{(k)}\sum_{j=\alpha(i)}^{2^{k-1}-1}\Delta L_j^{(k-1)}\big)^2] + \mathbb{E}[\big(\Delta W_{l,i}^{(k)} z_{\alpha(i)}^{(k-1)}(X_{2\alpha(i)}^{(k)})\cdot(\Delta W_i^{(k)} + \Delta W_{i+1}^{(k)})\big)^2] \\
&\quad + 2\mathbb{E}[(\Delta W_{l,i}^{(k)})^2 z_{\alpha(i)}^{(k-1)}(X_{2\alpha(i)}^{(k)})\cdot(\Delta W_i^{(k)} + \Delta W_{i+1}^{(k)})\sum_{j=\alpha(i)}^{2^{k-1}-1}\Delta L_j^{(k-1)}].
\end{aligned} \tag{3.3.26}$$

Since the Martingale increments $\Delta L_j^{(k-1)}$ are centered - i.e., $\mathbb{E}_j^{k-1}[\Delta L_j^{(k-1)}] = 0$ - it follows that the mixed terms $\mathbb{E}[(\Delta W_i^{(k)})^2\Delta L_{j_1}^{(k-1)}\Delta L_{j_2}^{(k-1)}]$ are equal to zero whenever $j_1 \ne j_2$, whence

$$\mathbb{E}[\big(\Delta W_{l,i}^{(k)}\sum_{j=\alpha(i)}^{2^{k-1}-1}\Delta L_j^{(k-1)}\big)^2] = \sum_{j=\alpha(i)}^{2^{k-1}-1}\mathbb{E}[\big(\Delta W_{l,i}^{(k)}\Delta L_j^{(k-1)}\big)^2].$$

Now, $(\Delta L_j^{(k-1)})^2$ equals $\Big(\sum_{r=1}^q \int_{t_j^{(k-1)}}^{t_{j+1}^{(k-1)}} (\tilde{z}_{r,t} - z_{r,j}^{(k-1)})dW_{r,t}\Big)^2$ where $W_{r,t}$ is the r-th component of W_t. Itô's Lemma implies that $\mathbb{E}_j^{k-1}[\int_{t_j^{(k-1)}}^{t_{j+1}^{(k-1)}} (\tilde{z}_{r_1,t} - z_{r_1,j}^{(k-1)})dW_{r_1,t} \int_{t_j^{(k-1)}}^{t_{j+1}^{(k-1)}} (\tilde{z}_{r_2,t} - z_{r_2,j}^{(k-1)})dW_{r_2,t}]$ is equal to $\mathbb{E}_j^{k-1}[\int_{t_j^{(k-1)}}^{t_{j+1}^{(k-1)}} (\tilde{z}_{r_1,t} - z_{r_1,j}^{(k-1)})(\tilde{z}_{r_2,t} - z_{r_2,j}^{(k-1)})d < W_{r_1,\cdot}, W_{r_2,\cdot} >_s]$, where $< W_{r_1,\cdot}, W_{r_2,\cdot} >_t$ is the quadratic covariation process. Since the components of the Brownian motion are independent, $< W_{r_1,\cdot}, W_{r_2,\cdot} >_t = 0$ whenever $r_1 \neq r_2$, and therefore it follows, from the Tower Law of conditional expectation, that

$$\mathbb{E}\big[\big(\Delta W_{l,i}^{(k)} \Delta L_j^{(k-1)}\big)^2\big] = \mathbb{E}[(\Delta W_{l,i}^{(k)})^2 \mathbb{E}_j^{(k,*)}[(\Delta L_j^{(k-1)})^2]] = \mathbb{E}[(\Delta W_{l,i}^{(k)})^2 \int_{t_j^{(k-1)}}^{t_{j+1}^{(k-1)}} |\tilde{z}_t - z_j^{(k-1)}|^2 dt].$$

for all $j > \alpha(i)$. To bound the above term, we decompose $\Delta W_{l,i}^{(k)}$ into a bounded part $\phi_{R\Delta^{(k)}}(\Delta W_{l,i}^{(k)})$ and an unbounded part $\Delta W_{l,i}^{(k)} - \phi_{R\Delta^{(k)}}(\Delta W_{l,i}^{(k)})$, just as in (3.3.23). To use this decomposition, one requires that $\mathbb{E}_j^{k-1}[\int_{t_j^{(k-1)}}^{t_{j+1}^{(k-1)}} |\tilde{z}_t - z_j^{(k-1)}|^2 dt]$ be bounded. Recalling the terminology $\Delta^{(k-1)} \tilde{z}_j^{(k-1)} := \mathbb{E}_j^{k-1}[\int_{t_j^{(k-1)}}^{t_{j+1}^{(k-1)}} z_t dt]$ from Section 3.2.3, the terms $\int_{t_j^{(k-1)}}^{t_{j+1}^{(k-1)}} |\tilde{z}_t - z_j^{(k-1)}|^2 dt$ are bounded by

$$\int_{t_j^{(k-1)}}^{t_{j+1}^{(k-1)}} |z_t - \tilde{z}_t|^2 dt + \int_{t_j^{(k-1)}}^{t_{j+1}^{(k-1)}} |z_t - \tilde{z}_j^{(k-1)}|^2 dt + |\tilde{z}_j^{(k-1)} - z_j^{(k-1)}|^2 \Delta^{(k-1)}. \tag{3.3.27}$$

Now, $|z_t|$ is bounded by C for all $t \in [0, T)$, as shown in Lemma 3.2.2, so $|\tilde{z}_j^{(k-1)}|$ is bounded by C and therefore $\int_{t_j^{(k-1)}}^{t_{j+1}^{(k-1)}} |z_t - \tilde{z}_j^{(k-1)}|^2 dt$ is bounded by $C\Delta^{(k-1)}$. Also, $|z_j^{(k-1)}|$ is bounded by C, as shown in Lemma 3.2.6, so $|\tilde{z}_j^{(k-1)} - z_j^{(k-1)}|^2 \Delta^{(k-1)}$ is bounded by $C\Delta^{(k-1)}$. Finally, using that

$$y_{t_j^{(k-1)}} - \tilde{y}_{t_j^{(k-1)}} + \int_{t_j^{(k-1)}}^T (z_t - \tilde{z}_t) dW_t = \Phi(X_T) - \Phi(X_{2^{k-1}}^{(k-1)})$$

$\mathbb{E}_j^{k-1}\big[\int_{t_j^{(k-1)}}^{t_{j+1}^{(k-1)}} |z_t - \tilde{z}_t|^2 dt\big]$ is bounded by $\mathbb{E}^{k-1}[|\Phi(X_T) - \Phi(X_{2^{k-1}}^{(k-1)})|^2] \leq C\Delta^{(k-1)}$ by (3.1.8). Therefore, using the bounds (3.3.22) and the results of (3.2.7) and that $\Delta^{(k-1)} = 2\Delta^{(k)}$,

$$\begin{aligned}\mathbb{E}[\big(\Delta W_{l,i}^{(k)} \sum_{j=\alpha(i)+1}^{2^{k-1}-1} \Delta L_j^{(k-1)}\big)^2] &\leq C\Delta^{(k)} \mathbb{E}[|\mathcal{N} - \phi_R(\mathcal{N})|^2] + CR\Delta^{(k)} \sum_{j=\alpha(i)+1}^{2^{k-1}-1} \int_{t_j^{(k-1)}}^{t_{j+1}^{(k-1)}} \mathbb{E}[|z_t - \tilde{z}_j^{(k-1)}|^2] dt \\ &\leq C\Delta^{(k)}(e^{-R/2} + R\Delta^{(k)}).\end{aligned} \tag{3.3.28}$$

It remains to bound the $\mathbb{E}[(\Delta W_{l,i}^{(k)} \Delta L_{\alpha(i)}^{(k-1)})^2]$. Using the Cauchy-Schwarz inequality, it follows that $\mathbb{E}[(\Delta W_{l,i}^{(k)} \Delta L_{\alpha(i)}^{(k-1)})^2] \leq \Delta^{(k)}\big(\mathbb{E}[(\Delta L_{\alpha(i)}^{(k-1)})^4]\big)^{1/2}$. The Burkholder-Davis-Gundy inequality then yields $\mathbb{E}[(\Delta L_{\alpha(i)}^{(k-1)})^4] \leq C\mathbb{E}[\big(\int_{t_j^{(k-1)}}^{t_{j+1}^{(k-1)}} |\tilde{z}_t - z_j^{(k-1)}|^2 dt\big)^2]$. Therefore, using the above bounds (3.3.27) on $\int_{t_j^{(k-1)}}^{t_{j+1}^{(k-1)}} |\tilde{z}_t - z_j^{(k-1)}|^2 dt$, $\mathbb{E}[(\Delta L_{\alpha(i)}^{(k-1)})^4]$ is bounded by $C(\Delta^{(k-1)})^2 + C\mathbb{E}[\big(\int_{t_{\alpha(i)}^{(k-1)}}^{t_{\alpha(i)+1}^{(k-1)}} |z_t - \tilde{z}_t|^2 dt\big)^2]$. Finally, [BDH+03, Proposition 3.2] gives precisely the result $\mathbb{E}[\big(\int_{t_{\alpha(i)}^{(k-1)}}^{t_{\alpha(i)+1}^{(k-1)}} |z_t - \tilde{z}_t|^2 dt\big)^2] \leq$

$C\mathbb{E}[|\Phi(X_T) - \Phi(X^{(k-1)}_{2^{k-1}})|^4]$. These bounds together with (3.1.8) yield

$$\mathbb{E}[(\Delta W^{(k)}_{l,i} \Delta L^{(k-1)}_{\alpha(i)})^2] \leq C(\Delta^{(k)})^2. \tag{3.3.29}$$

Substituting (3.3.28) and (3.3.29) into (3.3.26) yields

$$\begin{aligned}\mathcal{R}^{(2,M)}_{l,i} &\leq C\Delta^{(k)}(e^{-R/2} + R\Delta^{(k)}) + \mathbb{E}[\big(\Delta W^{(k)}_{l,i} z^{(k-1)}_{\alpha(i)}(X^{(k)}_{2\alpha(i)})(\Delta W^{(k)}_i + \Delta W^{(k)}_{i+1})\big)^2] \\ &\quad + 2\mathbb{E}[(\Delta W^{(k)}_{l,i})^2 z^{(k-1)}_{\alpha(i)}(X^{(k)}_{2\alpha(i)})\Delta W^{(k-1)}_{\alpha(i)} \sum_{j=\alpha(i)}^{2^{k-1}-1} \Delta L^{(k-1)}_j].\end{aligned}$$

Using that $|z^{(k-1)}_{\alpha(i)}(X^{(k)}_{2\alpha(i)})|$ is almost surely bounded by C, shown in Lemma 3.2.6, and the independence of the increments and components of the Brownian motion,

$$\mathbb{E}[\big(\Delta W^{(k)}_{l,i} z^{(k-1)}_{\alpha(i)}(X^{(k)}_{2\alpha(i)})(\Delta W^{(k)}_i + \Delta W^{(k)}_{i+1})\big)^2] \leq C(\Delta^{(k)})^2.$$

Using the Tower Law of conditional expectation and $\mathbb{E}^{(k,*)}_j[\Delta L^{(k-1)}_j] = 0$,

$$\mathbb{E}[(\Delta W^{(k)}_{l,i})^2 z^{(k-1)}_{\alpha(i)}(X^{(k)}_{2\alpha(i)})\Delta W^{(k-1)}_{\alpha(i)} \sum_{j=\alpha(i)}^{2^{k-1}-1} \Delta L^{(k-1)}_j] = \mathbb{E}[(\Delta W^{(k)}_{l,i})^2 z^{(k-1)}_{\alpha(i)}(X^{(k)}_{2\alpha(i)})\Delta W^{(k-1)}_{\alpha(i)} \Delta L^{(k-1)}_{\alpha(i)}].$$

By the Cauchy-Schwarz inequality and the almost sure bounds on $|z^{(k-1)}_{\alpha(i)}(X^{(k)}_{2\alpha(i)})|$, it follows that $|\mathbb{E}[(\Delta W^{(k)}_{l,i})^2 z^{(k-1)}_{\alpha(i)}(X^{(k)}_{2\alpha(i)})\Delta W^{(k-1)}_{\alpha(i)} \Delta L^{(k-1)}_{\alpha(i)}]|$ is bounded by $C\Delta^{(k)}\big(\mathbb{E}[(\Delta W^{(k-1)}_{\alpha(i)} \Delta L^{(k-1)}_{\alpha(i)})^2]\big)^{1/2}$. One applies the same techniques as in (3.3.29) to show that $\mathbb{E}[(\Delta W^{(k-1)}_{\alpha(i)} \Delta L^{(k-1)}_{\alpha(i)})^2]$ is bounded by $C(\Delta^{(k)})^2$, whence

$$|\mathbb{E}[(\Delta W^{(k)}_{l,i})^2 z^{(k-1)}_{\alpha(i)}(X^{(k)}_{2\alpha(i)}) \cdot (\Delta W^{(k)}_i + \Delta W^{(k)}_{i+1})\Delta L^{(k-1)}_{\alpha(i)}]| \leq C(\Delta^{(k)})^2.$$

The final bound for $\mathcal{R}^{(2,M)}_{l,i}$ is therefore $C\Delta^{(k)}(e^{-R/2} + R\Delta^{(k)})$.

To complete the proof, one takes the expectation $\mathbb{E}$ in (3.3.17), substitutes (3.3.18) and the bounds (3.3.24) on $\mathcal{R}^{(1,M)}_{l,i}$ and $C\Delta^{(k)}(e^{-R/2} + R\Delta^{(k)})$ on $\mathcal{R}^{(2,M)}_{l,i}$, and sums over $l = 1, \dots, q$. $\square$

3.3.3 Complexity analysis

Throughout this section, we will denote 2^κ by N.

In the previous section, we determined explicit, non-asymtotic upper bounds, given in Proposition 3.3.8 and Theorem 3.3.9, for the local error terms $\mathcal{E}(Y, k, i)$ and $\mathcal{E}(Z, k, i)$ - defined in (3.3.8) - in terms the basis functions, the number of simulations, and the free parameters R, $\varepsilon^{(Y)}_k$ and $\varepsilon^{(Z)}_k$. In this section, we use the master equations (3.3.15) in Proposition 3.3.8 and (3.3.16) in Theorem 3.3.9 to select specific numerical parameters so that the global error

$$\mathcal{E}(N) := \max_{0 \leq i < 2^\kappa} \mathcal{E}(Y, \kappa, i) + \sum_{i=0}^{2^\kappa - 1} \mathcal{E}(Z, \kappa, i)\Delta^{(\kappa)} \tag{3.3.30}$$

is bounded above by CN^{-1}, where $N = T(\Delta^{(\kappa)})^{-1}$ is the number of points in the time-grid $\pi^{(\kappa)}$. The rate of convergence is the same as the rate of *strong* convergence of the Euler scheme.

Before setting the specific numerical parameters, we use the master equations (3.3.15) and (3.3.16) with $k = \kappa$ to chose appropriate criteria to ensure (3.3.30). We see from the master equation (3.3.15) and the bound (3.3.14) on $T_{2,i}^{(Y,\kappa)}$ in Proposition 3.3.7, that it is sufficient to choose the numerical parameters so that, for all $i \in \{0, \ldots, 2^\kappa - 1\}$,

$$\max_{0 \leq i \leq 2^\kappa - 1} \{T_{1,i}^{(Y,\kappa)} + \frac{CK_{0,i}^{(\kappa)}}{M_\kappa}\} \leq CN^{-1}. \tag{3.3.31}$$

The master equation (3.3.16) shows that, for the z-part, it is sufficient to chose numerical parameters such that, for all $i \in \{0, \ldots, 2^\kappa - 1\}$,

$$\sum_{i=0}^{2^\kappa - 1} T_{1,i}^{(Z,\kappa)} \Delta^{(\kappa)} \leq C\Delta^{(\kappa)}, \quad \frac{R \sum_{i=0}^{2^\kappa - 1} \sum_{l=1}^{q} K_{l,i}^{(\kappa)} \Delta^{(\kappa)}}{M_\kappa} \leq C\Delta^{(\kappa)}, \tag{3.3.32}$$

$$\varepsilon_Y^{(\kappa-1)} + \max_{0 \leq j < 2^{\kappa-1}} \{T_{1,\alpha(i)}^{(Y,\kappa-1)} + T_{2,\alpha(i)}^{(Y,\kappa-1)}\} + e^{-R/2} + R\varepsilon_Z^{(\kappa-1)} \leq C\Delta^{(\kappa)}, \tag{3.3.33}$$

$$\exp\big(-CK_{0,\alpha(i)}^{(\kappa-1)} \ln(\varepsilon_Y^{(\kappa-1)}) - CM_{\kappa-1}\varepsilon_Y^{(\kappa-1)}\big) \leq C\Delta^{(\kappa)}, \tag{3.3.34}$$

$$R\sum_{l=1}^{q} \exp\big(-CK_{l,\alpha(i)}^{(\kappa-1)} \ln\big(\varepsilon_Z^{(\kappa-1)}\big) - CM_{\kappa-1}\varepsilon_Z^{(\kappa-1)}\big) \leq C\Delta^{(\kappa)} \tag{3.3.35}$$

$$R \sum_{j=0}^{2^{\kappa-1}-1} \sum_{r=1}^{q} \mathbb{E}[\|z_{r,j}^{(\kappa-1)} - z_{r,\alpha(i)}^{(\kappa-1,M)}\|_{k-1,\alpha(i),M}^2] \Delta^{(\kappa-1)} \leq C\Delta^{(\kappa)} \tag{3.3.36}$$

From equation (3.3.33), we set $R = C\log(N)$ in order to satisfy $e^{-R} \leq C\Delta^{(\kappa)} = CN^{-1}$. We will use (3.3.31) and (3.3.32) to calibrate the basis, the number of simulations and free parameters $\varepsilon_Y^{(\kappa)}$ and $\varepsilon_Z^{(\kappa)}$ in level κ. On the other hand, we use (3.3.33)-(3.3.36) to calibrate the number of simulations and the free parameters $\varepsilon_Y^{(\kappa-1)}$ and $\varepsilon_Z^{(\kappa-1)}$ in level $\kappa - 1$. In fact, we can generalize the criteria for the level $\kappa - 2$ to any level $k < \kappa$ by iteration with the master equation (3.3.16): for all $k \in \{0, \ldots, \kappa - 1\}$ and $i \in \{0, \ldots, 2^k - 1\}$,

$$\left.\begin{array}{l} \sum_{i=0}^{2^k-1} T_{1,i}^{(Z,k)} \Delta^{(k)} \leq C\Delta^{(k+1)}, \quad \big(R\sum_{i=0}^{2^k-1} \sum_{l=1}^{q} K_{l,i}^{(k)} \Delta^{(\kappa)}\big)/M_k \leq C\Delta^{(k+1)}, \\ \varepsilon_Y^{(k)} + \max_{0 \leq i < 2^k} \{T_{1,i}^{(Y,k)} + \frac{K_{0,i}^{(Y,k)}}{M_k}\} + R\varepsilon_Z^{(k-1)} \leq C\Delta^{(k+1)}, \\ \exp\big(-CK_{0,i}^{(k)} \ln(\varepsilon_Y^{(k)}) - CM_k\varepsilon_Y^{(k)}\big) \leq C\Delta^{(k+1)}, \\ R\sum_{l=1}^{q} \exp\big(-CK_{l,i}^{(k)} \ln\big(\varepsilon_Z^{(k)}\big) - CM_k\varepsilon_Z^{(k)}\big) \leq C\Delta^{(k+1)}. \end{array}\right\} \tag{3.3.37}$$

We begin by calibrating the basis to ensure that $\sum_{i=0}^{2^\kappa-1}\{T_{1,i}^{(Y,\kappa)} + T_{1,i}^{(Z,\kappa)}\}\Delta^{(\kappa)} \leq CN^{-1}$. Using Lemma 3.2.7, $\sum_{i=0}^{2^\kappa-1}\{T_{1,i}^{(Y,\kappa)} + T_{1,i}^{(Z,\kappa)}\}\Delta^{(\kappa)}$ is bounded above by

$$C\Delta^{(\kappa)} + \sum_{i=0}^{2^\kappa-1} \inf_{\alpha \in \mathbb{R}^{K_{0,i}^{(\kappa)}}} \mathbb{E}[|y_{t_i^{(\kappa)}}^{(\kappa)}(X_{t_i^{(\kappa)}}) - \alpha \cdot p_{0,i}^{(\kappa)}(X_{t_i^{(\kappa)}})|^2]\Delta^{(\kappa)}$$
$$+ \sum_{i=0}^{2^\kappa-1} \sum_{l=1}^{q} \inf_{\alpha \in \mathbb{R}^{K_{l,i}^{(\kappa)}}} \mathbb{E}[|z_{l,t_i^{(\kappa)}}^{(\kappa)}(X_{t_i^{(\kappa)}}) - \alpha \cdot p_{l,i}^{(\kappa)}(X_{t_i^{(\kappa)}})|^2]\Delta^{(\kappa)}$$

To perform the complexity analysis, we make the following assumption on the use of a local polynomial basis:

($\mathbf{A}_{\mathrm{poly}}$) Setting $\{p_{0,i}^{(\kappa,u)} : u = 1, \dots, K_{0,i}^{(\kappa)}\}$ to be local polynomials up to degree n on the hypercube partition $\{\mathcal{H}_u : u = 1, \dots, K_{0,i}^{(\kappa)}/n\}$ (with side-length $\delta_{y,i} = n^{1/d}\bar{R}/(K_{0,i}^{(\kappa)})^{1/d}$) on $[-\bar{R}, \bar{R}]^d$ and 0 outside $[\bar{R}, \bar{R}]^d$, the conclusion of Theorem 3.3.9 are valid. The constants C in Definition 3.1.1 may depend additionally on n.

Although this assumption may be rather strong, we wish to use local polynomials in order to investigate the effect of enhanced differentiability on the efficiency of the multilevel scheme.

Proposition 3.3.10. *Let $n \in \{2, 3, \dots\}$ and assume ($\mathbf{A_{diff}}$) and ($\mathbf{A}_{poly}$). There is a choice of basis functions and simulation sizes $(M_0, \dots, M_\kappa)$ and a constant C (independent of N) with dependencies as in ($\mathbf{A_{diff}}$) so that, for every N sufficiently large, the error $\mathcal{E}(N)$ given in (3.3.30) is bounded above by CN^{-1} for an algorithm complexity bounded above by*

$$\mathcal{C} = CN^{\frac{d}{2(n-1)} + \frac{d}{2} + 2} \log^{d+2}(N).$$

Proof. The functions $x \mapsto y_{t_i^{(\kappa)}}(x)$ (resp. $x \mapsto z_{t_i^{(\kappa)}}(x)$), $i \in \{0, \dots, 2^\kappa - 1\}$, are differentiable up to order n (resp. $n-1$) and the derivatives are bounded as in (3.2.6). Setting $\{p_{0,i}^{(\kappa,u)} : u = 1, \dots, K_{0,i}^{(\kappa)}\}$ to be local polynomials up to degree n on the hypercube partition $\{\mathcal{H}_u : u = 1, \dots, K_{0,i}^{(\kappa)}/n\}$ (with side-length $\delta_{y,i} = n^{1/d}\bar{R}/(K_{0,i}^{(\kappa)})^{1/d}$) on $[-\bar{R}, \bar{R}]^d$ and 0 outside $[\bar{R}, \bar{R}]^d$, we use Taylor's expansion to show that

$$\inf_{\alpha \in \mathbb{R}^{K_{0,i}^{(k)}}} \mathbb{E}[|y_{t_i^{(\kappa)}}(X_{t_i^{(\kappa)}}) - \alpha \cdot p_{0,i}^{(k)}(X_{t_i^{(\kappa)}})|^2] \le C\mathbb{P}(|X_{t_i^{(\kappa)}}|_\infty \ge \bar{R}) + \frac{C(\delta_{y,i})^{2n}}{(T - t_i^{(\kappa)})^{n-1}} \sum_{u=1}^{K_{0,i}^{(\kappa)}/n} \mathbb{P}(X_{t_i^{(\kappa)}} \in \mathcal{H}_u)$$

X has exponential moments thanks to ($\mathbf{A}_{\mathbf{b},\sigma}$), therefore $\mathbb{P}(|X_{t_i^{(\kappa)}}|_\infty \ge \bar{R}) \le Ce^{-C\bar{R}}$. Choosing $\bar{R} = C\log(N)$ and $\delta_{y,i} = CN^{-\frac{1}{2n}}(T - t_i^{(\kappa)})^{(1-\frac{1}{n})/2}\log(N)$ is therefore sufficient to ensure that $T_{1,i}^{(Y,\kappa)} \le CN^{-1}$, whence

$$K_{0,i}^{(\kappa)} = CN^{\frac{d}{2n}}(T - t_i^{(\kappa)})^{-\frac{d}{2}(1-\frac{1}{n})}\log^d(N).$$

This implies that $\|p_{0,i}^{(\kappa)}\|_\infty \le \bar{R}^n = C\log(N)^n$, but we normalize to reduce this upper bound to 1. Likewise, $\inf_{\alpha \in \mathbb{R}^{K_{l,i}^{(k)}}} \mathbb{E}[|z_{l,t_i^{(\kappa)}}(X_{t_i^{(\kappa)}}) - \alpha \cdot p_{l,i}^{(k)}(X_{t_i^{(\kappa)}})|^2] \le CN^{-1}$ for the basis choice of local polynomial up to degree $n-1$ on hypercubes with sides of length $\delta_{z,i} = CN^{\frac{1}{2(n-1)}}(T - t_i^{(\kappa)})^{1/2}\log(N)$, whence

$$K_{l,i}^{(\kappa)} = CN^{\frac{d}{2(n-1)}}(T - t_i^{(\kappa)})^{-d/2}\log^d(N),$$

on $[-\bar{R}, \bar{R}]^d$ and 0 outside $[-\bar{R}, \bar{R}]^d$. Here, $\|p_{l,i}^{(\kappa)}\|_\infty \le \bar{R}^{n-1} = \log(N)^d$.

Due to the selection of the basis, $T_{1,j}^{(Y,k)} + T_{1,j}^{(Z,k)} = T_{1,i}^{(Y,\kappa)} + T_{1,i}^{(Z,\kappa)}$ for all $k \in \{0, \dots, \kappa\}$ and $j \in \{0, \dots, 2^k - 1\}$ such that $t_j^{(k)} = t_i^{(\kappa)}$, Therefore $\sum_{j=0}^{2^k-1}\{T_{1,j}^{(Y,k)} + T_{1,j}^{(Z,k)}\}\Delta^{(k)} \le CN^{-1}$ for all $k \in \{0, \dots, \kappa\}$.

Equations (3.3.31) and (3.3.32) show that it is sufficient to take M_κ greater than

$$CN\Big(\max_{0\le i<2^\kappa} K_{0,i}^{(\kappa)} + R\sum_{i=0}^{2^\kappa-1}\sum_{l=0}^{q} K_{l,i}^{(\kappa)}\Delta^{(\kappa)}\Big).$$

We use an upper bound of this to calibrate M_κ. Using $(T-t_i^{(\kappa)})^{-\alpha} = T^{-\alpha}N^\alpha(N-i)^{-\alpha}$ for any $\alpha\ge 0$, it follows that $\max_{0\le i<2^\kappa} K_{0,i}^{(\kappa)} = CN^{d/2}\log^d(N)$ and $K_{l,i}^{(\kappa)} = CN^{\frac{dn}{2(n-1)}}(N-i)^{-d/2}\log^d(N)$. Let us determine an upper bound for the sum. Using the bounds $R\le C\log(N)$, and $N\Delta^{(\kappa)}\le C$, the sum is bounded above by

$$\begin{aligned} CN^{\frac{dn}{2(n-1)}}(1+\sum_{i=0}^{2^\kappa-2}(N-i)^{-d/2})\log^{d+1}(N) &\le CN^{\frac{dn}{2(n-1)}}(1+\int_0^{N-1}(N-t)^{-d/2}dt)\log^{d+1}(N)\\ &\le \begin{cases} CN^{\frac{dn}{2(n-1)}}\log^{d+2}(N) & \text{if } d\ge 2,\\ CN^{1+\frac{d}{2(n-1)}}\log^{d+1}(N) & \text{if } d<2.\end{cases} \end{aligned} \tag{3.3.38}$$

It is therefore sufficient to set M_κ equal to

$$\begin{array}{ll} C(N^{\frac{dn}{2(n-1)}}+N^{1+d/2})\log^{d+2}(N) & \text{if } d\ge 2,\\ C(N^{1+\frac{d}{2(n-1)}}+N^{1+d/2})\log^{1+d}(N) & \text{if } d=1.\end{array} \tag{3.3.39}$$

Next, we use (3.3.37) to determine the numerical parameters for the levels $k\in\{0,\dots,\kappa-1\}$. Firstly, we see that we need $\varepsilon_Y^{(k)}+R\varepsilon_Z^{(k)} = C\Delta^{(k+1)}$. There are two values that the number of simulations M_k must bound from above: the value

$$\max_{0\le i<2^k}\frac{K_{0,i}^{(k)}}{\Delta^{(k+1)}} + R\sum_{i=0}^{2^k-1}\sum_{l=1}^{q} K_{l,i}^{(k)}\frac{\Delta^{(k)}}{\Delta^{(k+1)}}$$

from the variance terms; and $C\log(2^{k+1})R\max_{0\le i<0}\max_{0\le l\le q} K_{l,i}^{(k)}/\Delta^{(k+1)}$ from the exponential terms. As before, but now writing $(T-t_i^{(k)})^{-\alpha} = T^{-\alpha}2^{k\alpha}(2^k-i)^{-\alpha}$ for any $\alpha\ge 0$,

$$\max_{0\le i<2^k} K_{0,i}^{(k)} = CN^{\frac{d}{2n}}2^{\frac{kd(n-1)}{2n}}\log^d(N),\quad \max_{0\le i<2^k}\max_{0\le l\le q} K_{l,i}^{(k)} = CN^{\frac{d}{2(n-1)}}2^{\frac{kd}{2}}\log^d(N).$$

Since $K_{l,i}^{(k)} = CN^{\frac{d}{2(n-1)}}2^{kd/2}(2^k-i)^{-d/2}\log^d(N)$, the sum $\sum_{i=0}^{2^k-1}\sum_{l=1}^{q} K_{l,i}^{(k)}$ is bounded above by

$$CN^{\frac{d}{2(n-1)}}2^{kd/2}(1+\int_0^{2^k-1}(2^k-t)^{-d/2}dt)\log^d(N) \le \begin{cases} CN^{\frac{d}{2(n-1)}+\frac{kd}{2\kappa}}\log^{1+d}(N) & \text{if } d\ge 2,\\ CN^{\frac{d}{2(n-1)}+\frac{k}{\kappa}}\log^d(N) & \text{if } d=1\end{cases}$$

where we have used $2^k = N^{k/\kappa}$ in the penultimate inequality. The requirements on M_k from the exponential terms are higher than those of the variance terms, and we choose

$$M_k = CN^{\frac{d}{2(n-1)}+\frac{kd}{2\kappa}+\frac{k}{\kappa}}\log^{d+2}(N) \tag{3.3.40}$$

We now come to the computation of the complexity of the algorithm. There are three main contributions to the complexity: the generation of the samples, sorting the samples into the hypercubes, and the computation of regression coefficients of the polynomials by least-squares regression within each hypercube. For each level $k \in \{0, \ldots, \kappa\}$ and each time point $i \in \{0, \ldots, 2^k\}$, computing the polynomial regression coefficients costs CM_k flops [GVL96, Section 5.5.9]; the sorting algorithm costs $CM_k \sum_{l=0}^{q} \log(K_{l,i}^{(k)}) = CM_k \log(N)$ flops [GL06, Section 1.4.1]. The total cost of these two operations is therefore dominated $C \log(N) \sum_{k=0}^{\kappa} 2^k M_k$. The generation of the samples costs $CN \sum_{k=0}^{\kappa} M_k$ flops. Using (3.3.40) and the formula $\sum_{k=0}^{\kappa-1} N^{\alpha k} = \frac{N^{\alpha\kappa}-1}{N^{\alpha}-1}$, it follows that the complexity is dominated by

$$\mathcal{C} = CN^{\frac{d}{2(n-1)}+\frac{d}{2}+2} \log^{d+2}(N).$$

$\square$

Proposition 3.3.11. *Let $n \in \{2, 3, \ldots\}$ and assume $(\mathbf{A}_{\partial\mathbf{\Phi}})$ and $(\mathbf{A}_{poly})$. There is a choice of basis functions and simulation sizes $(M_0, \ldots, M_\kappa)$ and a constant C (independent of N) as in $(\mathbf{A}_{\partial\mathbf{\Phi}})$ so that, for every N sufficiently large, the error $\mathcal{E}(N)$ given in* (3.3.30) *is bounded above by CN^{-1} for an algorithm complexity bounded above by*

$$\mathcal{C} = CN^{2+\frac{d}{2(n-1)}} \log^{d+1}(N).$$

Proof. By Lemma 3.2.3, the derivatives of $y_{t_i^{(\kappa)}}(\cdot)$ and $z_{t_i^{(\kappa)}}(\cdot)$ are uniformly bounded in time and i. We use a local polynomials on hypercubes basis as in Proposition 3.3.10, but we calibrate the hypercube lengths according to the improved gradient bounds of Lemma 3.2.3. Using the Taylor expansion to calibrate the length of the hypercubes, used as above, this implies that $\delta_{y,i} = CN^{-1/(2n)}$ and $\delta_{z,i} = CN^{-1/\big(2(n-1)\big)}$ for all $i \in \{0, \ldots, N-1\}$. The bases will be of size $K_{0,i}^{(k)} = CN^{d/(2n)} \log^d(N)$ and $K_{0,i}^{(k)} = CN^{d/\big(2(n-1)\big)} \log^d(N)$ for all $k \in \{0, \ldots, \kappa\}$, $i \in \{0, \ldots, N-1\}$ and $l \in \{1, \ldots, q\}$. Following the calibration procedure of Proposition 3.3.10 above, we set the sample sizes

$$M_\kappa = CN^{1+d/\big(2(n-1)\big)} \log^{1+d}(N), \quad M_k = CN^{k/\kappa+d/\big(2(n-1)\big)} \log^{1+d}(N)$$

for all $k \in \{0, \ldots, \kappa-1\}$. Therefore, the overall complexity is dominated by $\mathcal{C} = CN^{2+\frac{d}{2(n-1)}} \log^{d+1}(N)$. $\square$

The complexity of the algorithm under differentiable terminal condition $(\mathbf{A}_{\partial\mathbf{\Phi}})$ is clearly an improvement on that general Lipschitz case. The terms related to the dimension d in the exponent of N in Proposition 3.3.11 are decreasing to 0 as the degree of the local polynomials increases, whereas those of Proposition 3.3.10 converge to $d/2$. This occurs due to the improved bounds on $|\nabla_x^r y_{t_i^{(\kappa)}}(x)|_\infty$ and $|\nabla_x^r z_{t_i^{(\kappa)}}(x)|_\infty$, which allow one to increase the length of the hypercubes and thereby reduce the overall complexity of the algorithm.

At this point, it is important to remark that the bounds on $|\nabla_x^r y_{t_i^{(\kappa)}}(x)|_\infty$ obtained from Lemma 3.2.3 depend on $|\nabla_x^r \Phi(x)|_\infty$. This means that if one wishes to smooth $x \mapsto \Phi(x)$ to obtain the better algorithm complexity of Proposition 3.3.11 rather than Proposition 3.3.10, this must be done rather carefully. Finding efficient smoothing procedures to optimize the multilevel algorithm may be a fruitful direction of future research.

3.3.4 Parallel computing

Taking a closer look at (3.3.6), we observe that the random variable $\Xi_{l,k,i}^{(m)}$ is constructed using the function Φ, the approximations $y_{\alpha(i)}^{(k-1,M)}(x)$ and $z_j^{(k-1,M)}(x)$ ($j \in \{\alpha(i)+1, \ldots, 2^{k-1}-1\}$) and the samples $X_r^{(k,m)}$ ($r \in \{i, \ldots, 2^k\}$). Since all of these components are precomputed, we see that it is possible to spread the computation of the regression coefficients $\alpha_{l,i}^{(k,M)}$ for $l \in \{0, \ldots, q\}$ and $i \in \{0, \ldots, 2^k - 1\}$ over multiple processors.

Moreover, the computations of the paths of $(X_j^{(k,m)})_{0 \le j \le 2^k}$ can also be spread over different processors for each $k \in \{0, \ldots, \kappa\}$ and $m \in \{1, \ldots, M_k\}$.

Parallelizability is a strong advantage of this method.

3.3.5 Comparison to LSMDP

Let us recall the LSMDP scheme of Chapter 2. For this algorithm, we work only on the finest time-grid $\pi^{(\kappa)}$, and require only the samples $(X^{(\kappa,1:M_\kappa)}, \Delta W^{(\kappa,1:M_\kappa)})$ associated to this level. We use the terminology of Section 3.3.1. Refering to Definition 3.3.1 and (3.3.4) for the definition of $LS(\cdot)$, let $\bar{\alpha}_{0,i}^{(M)}$ solve $LS(X_i^{(\kappa,1:M_\kappa)}, \Phi(X_N^{(\kappa,1:M_\kappa)}), p_{0,i}^{(\kappa)}(x))$, and, setting $\Delta^{(\kappa)}\Xi_{l,i}^{(m)} := \Phi(X_N^{(\kappa,m)})\Delta W_i^{(\kappa,m)}$, let $\bar{\alpha}_{l,i}^{(M)}$ be the coefficient solving $LS(X_i^{(\kappa,1:M_\kappa)}, \Xi_{l,i}^{(1,1:M_\kappa)}(X_N^{(\kappa,1:M_\kappa)}), p_{l,i}^{(\kappa)}(x))$. Using the bounds of Lemma 3.2.6, we set the approximating functions

$$y_i^{(M)}(x) := -C \vee \bar{\alpha}_{0,i}^{(M)} \cdot p_{0,i}^{(\kappa)}(x) \wedge C, \quad z_{l,i}^{(M)}(x) := -C \vee \bar{\alpha}_{l,i}^{(M)} \cdot p_{l,i}^{(\kappa)}(x) \wedge C$$

Note that this is also equivalent to the Bender-Denk scheme [BD07], in which there are no Picard iterations because the BSDE-driver is 0.

Using the same techniques as in Proposition 3.3.8 and Proposition 3.3.7, it follows that, for every $i \in \{0, \ldots, N-1\}$ and $l \in \{1, \ldots, q\}$, the local error terms have the following bounds:

$$\mathbb{E}[\|y_i^{(\kappa)} - y_i^{(M)}\|_{\kappa,i,M}^2] \le C \inf_{\alpha \in \mathbb{R}^{K_{0,i}^{(\kappa)}}} \mathbb{E}[|y_{t_i^{(\kappa)}}(X_i^{(\kappa)}) - \alpha \cdot p_{0,i}^{(\kappa)}(X_i^{(\kappa)})|^2] + \frac{CK_{0,i}^{(\kappa)}}{M_\kappa},$$

$$\mathbb{E}[\|z_{l,i}^{(\kappa)} - z_{l,i}^{(M)}\|_{\kappa,i,M}^2] \le C \inf_{\alpha \in \mathbb{R}^{K_{l,i}^{(\kappa)}}} \mathbb{E}[|z_{l,t_i^{(\kappa)}}(X_i^{(\kappa)}) - \alpha \cdot p_{l,i}^{(\kappa)}(X_i^{(\kappa)})|^2] + \frac{CK_{l,i}^{(\kappa)}}{\Delta^{(\kappa)} M_\kappa}.$$

Proposition 3.3.12. *Let $n \in \{2, 3, \ldots\}$ and assume $(\mathbf{A_{diff}})$. There is a choice of basis functions and simulation sizes M_κ and a constant C (independent of N) with dependencies as in $(\mathbf{A_{diff}})$ so that, for every N sufficiently large, the error of the LSMDP scheme is bounded above by CN^{-1} for an algorithm complexity bounded above by*

$$\mathcal{C} = C(N^{2+\frac{d}{2(n-1)}+\frac{d}{2}} \mathbf{1}_{d>2} + N^{3+\frac{d}{2(n-1)}} \mathbf{1}_{d \le 2}) \log^{1+d}(N).$$

Proof. Since the bias terms are the same as for the multilevel scheme in Section 3.3.3, we exactly the same local polynomial on hypercubes basis with the same hypercube lengths as in

Proposition 3.3.10. we select the number of simulations M_κ that dominates the expression

$$CN\{\max_{0\le i\le N-1} K^{(\kappa)}_{0,i} + N\sum_{i=0}^{N-1}\sum_{l=1}^{q} K^{(\kappa)}_{l,i}\Delta^{(\kappa)}\} \le CN^{1+\frac{dn}{2(n-1)}}\sum_{i=0}^{N-1}(N-i)^{-d/2}\log^d(N)$$
$$\le \begin{cases} CN^{1+\frac{d}{2(n-1)}+\frac{d}{2}}\log^{1+d}(N) & \text{if } d\ge 2,\\ CN^{2+\frac{d}{2(n-1)}}\log^d(N) & \text{if } d=1.\end{cases}$$

The overall complexity will be dominated by the cost of the sorting algorithm, which is $CNM_\kappa\log(N) = C(N^{2+\frac{d}{2(n-1)}+\frac{d}{2}}\mathbf{1}_{d\ge 2} + N^{3+\frac{d}{2(n-1)}}\mathbf{1}_{d=1})\log^{1+d}(N)$. □

Let us compare the results of Proposition 3.3.10 and Proposition 3.3.12. The exponents of N are the same for both the algorithms, suggesting that the algorithm complexity is very similar. If we look closely at the proofs, the bias due to the selection of the basis is the same for both algorithms, therefore the main difference in complexity is due to the selection of the number of simulations. In fact, the number of simulations for the multilevel scheme are essentially selected by dominating the largest basis size, $\max_i K^{(\kappa)}_{l,i}$, whereas the number of simulations for the LSMDP are selected by dominating the sum of the basis sizes, $\sum_i K^{(\kappa)}_{l,i}$. It is a basic fact - a statement of the equivalence of the $|\cdot|_\infty$ and $|\cdot|_1$ norms on $\mathbb{R}^N$ - that

$$\max_{0\le i\le N-1} K^{(\kappa)}_{l,i} \le \sum_{i=0}^{N-1} K^{(\kappa)}_{l,i} \le N\max_{0\le i\le N-1} K^{(\kappa)}_{l,i}. \tag{3.3.41}$$

The heuristic argument above combined with (3.3.41) shows us why the complexities of the two algorithms are so similar. Due to the basis selection, the sum of the basis sizes $\sum_i K^{(\kappa)}_{l,i}$ is of the same order as $\max_i K^{(\kappa)}_{l,i}$. The lower inequality of (3.3.41) suggests, however, that this is an extremal example: it should not be possible to obtain a better complexity for LSMDP (in terms of $O(N)$) than for the multilevel scheme using the methods we have presented above. On the other hand, the upper inequality in (3.3.41) suggests that it may be possible for the multilevel scheme to obtain an order 1 improvement of efficiency compared to LSMDP for certain classes of bases. This phenomenon is observed in a natural setting when $(\mathbf{A}_{\partial\boldsymbol{\Phi}})$ is in force.

Proposition 3.3.13. *Let $n\in\{2,3,\dots\}$ and assume $(\mathbf{A}_{\partial\boldsymbol{\Phi}})$. There is a choice of basis functions and simulation sizes M_κ and a constant C (independent of N) as in $(\mathbf{A}_{\partial\boldsymbol{\Phi}})$ so that, for every N sufficiently large, the error of the LSMDP scheme is bounded above by CN^{-1} for an algorithm complexity bounded above by*

$$\mathcal{C} = CN^{3+\frac{d}{2(n-1)}}\log^{d+1}(N).$$

Proof. We choose exactly the same basis as in Proposition 3.3.11, and set the number of simulations using the technique of Proposition 3.3.12 but using the basis sizes $K^{(\kappa)}_{0,i} = CN^{d/(2n)}\log^d(N)$ and $K^{(\kappa)}_{l,i} = K^{(\kappa)}_{0,i} = CN^{d/(2(n-1))}\log^d(N)$. The complexity is computed as in Proposition 3.3.12. □

Comparing Proposition 3.3.11 with Proposition 3.3.13, we see observe the order 1 improvement in the complexity predicted in the heuristics.

3.4 Zero terminal condition BSDE

3.4.1 Continuous time theory

We now consider the approximation of $(\bar{y}, \bar{z})$, the nonlinear BSDE in (3.1.1). The methods we use here can be equally well applied to (Y, Z), the full satisfying (3.1.4), and we do so to obtain a fair comparison. In this sense, we update the results of Chapter 2. In this section, we make use of the following equivalent definition of $(\bar{y}, \bar{z})$:

$$\bar{y}_t = \int_t^T \bar{f}(r, X_r, \bar{y}_r, \bar{z}_r)dr - \int_t^T \bar{z}_r dW_r, \quad \bar{f}(r, x, y, z) := f(r, x, y_r + y, z_r + z). \tag{3.4.1}$$

The driver function $\bar{f}(t, x, y, z)$ maps $[0, T] \times \Omega \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^q$ to $\mathbb{R}$, is $\mathcal{P} \otimes \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B} \otimes \mathcal{B}(\mathbb{R}^q)$-measurable, and is Lipschitz continuous in (x, y, z) uniformly in t due to the Lipschitz continuity of f:

$$\begin{aligned} |\bar{f}(t, x, y, z) - \bar{f}(t, x', y', z')| &= |f(t, x, y_t + y, z_t + z) - f(t, x', y_t + y', z_t + z')| \\ &\leq L_f(|x - x'| + |y - y'| + |z - z'|). \end{aligned}$$

The BSDE (3.4.1) has terminal condition 0.

The following lemma is a restatement of [EKPQ97, Theorem 4.1] which yields a Markov representation for the nonlinear BSDE $(\bar{y}, \bar{z})$. We make use of this result, so include it without proof for the convenience of the reader.

Lemma 3.4.1. *For all $t \in [0, T)$, there exists deterministic measurable functions $\bar{y}_t : \mathbb{R}^d \to \mathbb{R}$ and $\bar{z}_t : \mathbb{R}^d \to \mathbb{R}^q$ such that $\bar{y}_t = \bar{y}_t(X_t)$ and $\bar{z}_t = \bar{z}_t(X_t)$ almost surely.*

As in Section 3.2.2, we introduce smoothness assumptions for the functions $\bar{y}_t(\cdot)$ and $\bar{z}_t(\cdot)$. These assumptions will only be in force if explicitly stated.

$(\mathbf{A}'_{\mathbf{diff}})$ For some integer n greater than 2, the assumption $(\mathbf{A}_{\mathbf{diff}})$ is in force. Moreover, the functions $\bar{y}_t(\cdot)$ are n-times continuously differentiable, and $\bar{z}_t(\cdot)$ are $(n-1)$-times continuously differentiable, and there is a constant C as in $(\mathbf{A}_{\mathbf{diff}})$ such that

$$\left.\begin{array}{rcll} \left|\frac{\partial^r}{\partial x_{i_1} \ldots \partial x_{i_r}} \bar{y}_t(x)\right| & \leq & C(T-t)^{(2-r)/2}, & \text{for } r = 1, \ldots, n, \\ \left|\frac{\partial^r}{\partial x_{i_1} \ldots \partial x_{i_r}} \bar{z}_t(x)\right| & \leq & C(T-t)^{(1-r)/2}, & \text{for } r = 1, \ldots, n-1 \end{array}\right\} \text{ for all } t \in [0, T) \tag{3.4.2}$$

where $\{i_1, \ldots, i_r\} \in \{1, \ldots, d\}^r$.

The assumption $(\mathbf{A}'_{\mathbf{diff}})$ is quite natural. In fact, in the recent work of [CD12, Theorem 1.4], gradient bounds of the form (3.4.2) are obtained under assumptions that the coefficients b and σ of the SDE 3.1.2 are time-homogeneous and the driver f is n-times differentiable; it is also not necessary that σ satisfies $(\mathbf{A}_{\text{u.e.}})$, but a weaker condition. We do not use these results in our analysis, but feel it is useful to mention them to motivate the assumption $(\mathbf{A}'_{\mathbf{diff}})$. There is also an order improvement in the gradient bounds (3.2.6). This motivated by [GM10, Remark 2.2], where it is proven that $|\bar{z}_t(x)| \leq C(T-t)^{1/2}$. Given the relationship between $\bar{z}_t(x) = \sigma(t,x)^\top \nabla_x \bar{y}_t(x)$ given in [EKPQ97, Corollary 4.1], it seems natural to assume the improved convergence rate. Obtaining precise gradient bounds of the form (3.4.2) is a work in progress.

3.4.2 Discrete nonlinear BSDE

In this section, we only consider the discretization of $(\bar{y}, \bar{z})$ on the finest time-grid, $\pi^{(\kappa)}$. We simplify our notation by introducing $t_i := t_i^{(\kappa)}$, $\Delta := \Delta^{(\kappa)}$, $\Delta W_i := \Delta W_i^{(\kappa)}$, $X_i := X_i^{(\kappa)}$, and the conditional expectation operator $\mathbb{E}_i[\cdot] := \mathbb{E}[\cdot|\mathcal{F}_{t_i}^{(W)}]$. Also, we define the nonlinear functions $\bar{f}_j(x,y,z) := f(t_j, x, y_j^{(k)}(x) + y, z_j^{(k)}(x) + z)$ and $f_j(x,y,z) := f(t_j,x,y,z)$ for $j \in \{0,\dots,N-1\}$, and $\bar{f}_N(x,y,z) := f(T,x,\Phi(x)+y,0)$ and $f_N(x,y,z) = f(T,x,y,0)$.

For each $j \in \{1,\dots,N\}$, let $f_j : \Omega^{(W)} \times \mathbb{R} \times \mathbb{R}^q \to \mathbb{R}$ $(j = 1,\dots,N)$ be $\mathcal{F}^{(W)} \otimes \mathcal{B} \otimes \mathcal{B}(\mathbb{R}^q)$-measurable, $(y,z) \in \mathbb{R} \times \mathbb{R}^q$. Let $\xi \in \mathbf{L}_2(\mathcal{F}^{(W)})$. Consider the following discrete nonlinear BSDE:

Definition 3.4.2. *The discrete processes $(Y_i, Z_i)_{0 \leq i \leq N}$ are said to satisfy a discrete BSDE with driver f_j and terminal condition ξ if*

$$\left.\begin{array}{ll} Y_N & := \xi, \quad Z_N = 0, \\ Y_i & := \mathbb{E}_i[\xi] + \sum_{j=i+1}^N \mathbb{E}_i[f_j(Y_j, Z_j)]\Delta, \\ Z_i & := \frac{\mathbb{E}_i[\xi \Delta W_i]}{\Delta} + \sum_{j=i+1}^N \mathbb{E}_j[\Delta W_i f_j(Y_j, Z_j)]. \end{array}\right\} \tag{3.4.3}$$

Notice that the formulation of (3.4.3) is slightly different to the discrete BSDE (2.1.4) of Chapter 2: the driver function f_j takes a y-argument Y_j rather than Y_{j+1}, but the summation starts at $i+1$ rather than i to preserve the explicit nature of the formulation. We can use the same arguments as Proposition 2.3.3 of Chapter 2 to prove analogous a priori estimates.

Proposition 3.4.3. *Suppose that $(Y_{u,i}, Z_{u,i})_{0 \leq i \leq N}$ $(u = 1,2)$ solves a discrete BSDE with driver $f_{u,j}(y,z)$ and terminal condition ξ_u. Moreover, suppose that $(y,z) \mapsto f_{2,j}(y,z)$ is Lipschitz continuous with Lipschitz constant L_f uniformly in j, and that $f_{u,j}(Y_{u,j}, Z_{u,j}) \in \mathbf{L}_2(\mathcal{F}^{(W)})$.*

For all $i \in \{0,\dots,N\}$, set

$$\Delta Y_i := Y_{1,i} - Y_{2,i}, \quad \Delta Z_i := Z_{1,i} - Z_{2,i}, \quad \Delta\xi = \xi_1 - \xi_2,$$
$$\Delta f_j := f_{1,j}(Y_{1,j}, Z_{1,j}) - f_{2,j}(Y_{1,j}, Z_{1,j}), \quad j = i+1, \dots N.$$

Let Δ be sufficiently small and $\gamma \in (0,\infty)$ be such that $6q(\Delta + \frac{1}{\gamma})L_f^2 \leq 1$, and define by $\Gamma_i = (1+\gamma\Delta)^i$. Then for $\bar{C} := 2q + (1+T)e^{T/2}$,

$$|\Delta Y_i|^2 \Gamma_i + \sum_{j=i}^N \Gamma_j \mathbb{E}_i[|\Delta Z_j|^2]\Delta \leq \bar{C}\mathbb{E}_i[|\Delta\xi|^2] + 3\bar{C} \sum_{j=i+1}^N \Gamma_j (\frac{1}{\gamma} + \Delta)\mathbb{E}_i[|\Delta f_j|^2]\Delta$$

We now come to the discretization of the nonlinear BSDE $(\bar{y}, \bar{z})$.

Definition 3.4.4. *Let $(\bar{y}_i, \bar{z}_i)_{0 \leq i \leq N}$ be the discrete BSDE solving (3.4.3) with driver $f_j(\omega, y, z) = \bar{f}_j\big(X_j(\omega), y, z\big)$ and terminal condition $\xi \equiv 0$.*

Recall the definition of the Markov chains $X^{(i,x)}$ in Section 3.1.3, and define iteratively the functions

$$\left.\begin{array}{rl}\bar{y}_N(x) & := 0, \quad \bar{z}_N(x) := 0, \quad \bar{z}_i(x) := \mathbb{E}[\Delta W_i \bar{y}_{i+1}(X_{i+1}^{(i,x)})], \\ \bar{y}_i(x) & := \mathbb{E}[y_{i+1}(X_{i+1}^{(i,x)}) + \bar{f}_{i+1}\big(X_{i+1}^{(i,x)}, \bar{y}_{i+1}(X_{i+1}^{(i,x)}), \bar{z}_{i+1}(X_{i+1}^{(i,x)})\big)]\end{array}\right\} \quad \forall i \in \{0, \ldots, N-1\}. \tag{3.4.4}$$

Lemma 3.4.5. *For each $i \in \{0, \ldots, N\}$, $\bar{y}_i = \bar{y}_i(X_i)$ and $\bar{z}_i = \bar{z}_i(X_i)$ almost surely.*

Moreover, given $x \in \mathbb{R}^d$ and $i \in \{0, \ldots, N\}$, the discrete processes given by

$$Y_j := \bar{y}_j(X_j^{(i,x)})\mathbf{1}_{j \geq i} + \bar{y}_i(x)\mathbf{1}_{j<i}, \quad Z_j := \bar{z}_j(X_j^{(i,x)})\mathbf{1}_{j \geq i}$$

define a discrete BSDE with terminal condition 0 and driver $f_j(y,z) = \bar{f}_j(X_j^{(i,x)}, y, z)\mathbf{1}_{j \geq i}$.

Proof. Use Lemma 3.6.1 in the same way as in the proof of Lemma 3.2.5.

Proposition 3.4.6. *Assuming $\Delta \leq (12qL_f^2)^{-1}$, there is a constant C such that $|\bar{y}_i(x)| \leq C(T-t_i)$ and $|\bar{z}_i(x)| \leq C(T-t_i)/\Delta^{1/2}$ for all $i \in \{0, \ldots, N\}$ and $x \in \mathbb{R}^d$.*

Proof. Let $(Y_{1,j}, Z_{1,j}) = (0,0)$ and $(Y_{2,j}, Z_{2,j}) = \big(\bar{y}_j(X_j^{(i,x)}), \bar{z}_j(X_j^{(i,x)})\big)$ for all $j \in \{i, \ldots, N\}$. $(Y_{1,j}, Z_{1,j})$ is a discrete BSDE with driver and terminal condition 0, and $(Y_{2,j}, Z_{2,j})$ is a discrete BSDE with terminal condition 0 and driver $f_j(y,z) = \bar{f}_j(X_j^{(i,x)}, y, z)$ (Lemma 3.4.5). We obtain from Proposition 3.4.3, using $\gamma = 12qL_f^2$ and $\Gamma_j \leq e^{\gamma T}$ for all j, the uniform bounds of Lemma 3.2.6 on $|y_j^{(\kappa)}(x)|$ and $|z_j^{(\kappa)}(x)|$, and the Lipschitz continuity and uniform bounds of $f(t,x,y,z)$ from Assumption $(\mathbf{A_f})$, that

$$\begin{aligned}|\bar{y}_i(x)|^2 &\leq C \sum_{j=i+1}^{N} \mathbb{E}_i[|\bar{f}_j(X_j^{(i,x)}, 0, 0) - f_j(X_j^{(i,x)}, 0, 0)|^2]\Delta + C \sum_{j=i+1}^{N} |f_j(X_j^{(i,x)}, 0, 0)|^2 \Delta \\ &\leq C \sum_{j=i+1}^{N} L_f^2 \mathbb{E}_i[|y_j^{(\kappa)}(X_j^{(i,x)})|^2 + |z_j^{(\kappa)}(X_j^{(i,x)})|^2]\Delta + C(T-t_i) \leq C(T-t_i).\end{aligned}$$

To prove the bound on $\bar{z}_i$, let (Y_i, Z_i) be the discrete BSDE with terminal condition $\Phi(X_N)$ and driver $f_j(y,z) = f_j(X_j, y, z)$ and observe that, for all $i \in \{0, \ldots, N\}$, $\bar{y}_i = Y_i - y_i^{(\kappa)}$ and $\bar{z}_i = Z_i - z_i^{(\kappa)}$. One can show analogously to Corollary 2.3.11 of Chapter 2 that $|Z_i| \leq C$ almost surely. Additionally, the almost sure bound $|z_i^{(\kappa)}| \leq C$ is obtained from Lemma 3.2.2. Therefore, $|\bar{z}_i| \leq |Z_i| + |z_i^{(\kappa)}| \leq C$ almost surely.

Using these a priori bounds and the Lipschitz continuity and bounds of $(\mathbf{A_f})$ on $f(t,\cdot)$, it follows that there is a constant C such that $\bar{f}_i\big(x, \bar{y}_i(x), \bar{z}_i(x)\big) \leq C$ for all $x \in \mathbb{R}^d$ and $i \in \{0, \ldots, N\}$. The final result now follows from the triangle inequality:

$$\begin{aligned}|\bar{y}_i(x)| &\leq \sum_{j=i+1}^{N} \mathbb{E}_i[|\bar{f}_j\big(X_j^{(i,x)}, \bar{y}_j(X_j^{(i,x)}), \bar{z}_j(X_j^{(i,x)})\big)|]\Delta \leq C(T-t_i), \\ \Delta|\bar{z}_i(x)| &\leq \sum_{j=i+1}^{N} \mathbb{E}_i[|\Delta W_i \bar{f}_j\big(X_j^{(i,x)}, \bar{y}_j(X_j^{(i,x)}), \bar{z}_j(X_j^{(i,x)})\big)|]\Delta \leq C\mathbb{E}[|\Delta W_i|^2]^{1/2}(T-t_i).\end{aligned}$$

□

3.4.3 Monte Carlo scheme

As in Section 3.3, we approximate the functions $\bar{y}(\cdot)$ (resp. $\bar{z}(\cdot)$) by $\bar{y}_i^{(M)}(\cdot)$ (resp. $\bar{z}_i^{(M)}(\cdot)$) using least-square regression recursively to approximate $(\bar{y}_i^{(M)}(x), \bar{z}_i^{(M)}(x))$ for $i < N$. The recursion takes place over the time-points rather than over the multigrids: initializing with $\bar{y}_N^{(M)}(x) \equiv 0$ and $\bar{z}_N^{(M)}(x) \equiv 0$, the computation of functions $(\bar{y}_i^{(M)}(x), \bar{z}_i^{(M)}(x))$ for $i < N$ makes use of the functions $(\bar{y}_j^{(M)}(x), \bar{z}_j^{(M)}(x))_{i+1 \le j \le N}$. The terminology and techniques are similar to Section 3.3.1, and we make reference to this section where possible.

Basis functions. For each $l = 0, \ldots, q$ and $i = 0, \ldots 2^k - 1$, we are given basis functions $\{p_{r,l,i} : \mathbb{R}^d \to \mathbb{R} \ : \ 1 \le r \le K_{l,i}^{(k)}\}$. We write $p_{l,i}(x) := \big(p_{1,l,i}(x), \ldots, p_{K_{l,i}^{(k)},l,i}(x)\big)$ for the vector of basis functions. For every $r \in \{0, \ldots, K_{l,i}^{(k)}\}$, $p_{r,l,i}$ is measurable and $p_{r,l,i}(X_i)$ is in $\mathbf{L}_2(\mathcal{F}_i^{(W)})$.

Simulations. For $i = 0, \ldots, N-1$, let M_i be an integer and generate M_i copies $\{(\Omega^{(i,m)}, \mathcal{F}^{(i,m)}, \mathbb{P}^{(i,m)}) \ : \ m = 1, \ldots, M_i\}$ of the probability space $(\Omega^{(W)}, \mathcal{F}^{(W)}, \mathbb{P}^{(W)})$. Define $X^{(i,m)}$ to be the copy of X and $\Delta W^{(i,m)}$ to be the copy of ΔW in $(\Omega^{(i,m)}, \mathcal{F}^{(i,m)}, \mathbb{P}^{(i,m)})$; we call these objects the *simulations*, because, in practice, one generates these objects using random number generators. Define by $(\Omega^{(NL)}, \mathcal{F}^{(NL)}, \mathbb{P}^{(NL)})$ the product space of $(\Omega, \mathcal{F}, \mathbb{P})$ and $\bigotimes_{k,m}(\Omega^{(k,m)}, \mathcal{F}^{(k,m)}, \mathbb{P}^{(k,m)})$, and $\mathbb{E}^{(NL)}$ the associated expectation operator. The clouds $\{(X^{(i,1:M_i)}, \Delta W^{(i,1:M_i)}) \ : \ i = 0, \ldots, N-1\}$ are independent in $(\Omega^{(NL)}, \mathcal{F}^{(NL)}, \mathbb{P}^{(NL)})$. Moreover, they are independent of the simulations used for the multilevel algorithm of Section 3.3.1.

To reduce the notational complexity, we extend the expectation operator $\mathbb{E}$ and measure $\mathbb{P}$ to $(\Omega^{(NL)}, \mathcal{F}^{(NL)})$ in order to avoid writing $\mathbb{E}^{(NL)}$ and $\mathbb{P}^{(NL)}$: for any $\mathcal{F}^{(NL)}$-measurable random variable $\mathcal{Y}$, $\mathbb{E}[\mathcal{Y}] := \mathbb{E}^{(NL)}[\mathcal{Y}]$. Also, we extend the conditional expectation operator $\mathbb{E}_i$ to $(\Omega^{(NL)}, \mathcal{F}^{(NL)})$: $\mathbb{E}_i[\mathcal{Y}] := \mathbb{E}[\mathcal{Y}|\mathcal{F}_{t_i}^{(W)}]$.

Recall the definition of $\mathcal{F}^{(*)}$ in (3.3.5) and define the σ-algebras

$$\left.\begin{array}{rcl} \mathcal{F}_i^{(*)} & := & \mathcal{F}^{(*)} \vee \sigma(X_\nu^{(j,m)}, \Delta W_\nu^{(j,m)} \ : \ j > i \ ; \ \nu = 0, \ldots, N \ ; \ 1 \le m \le M_j), \\ \mathcal{F}_i^{(M)} & := & \mathcal{F}_i^{(*)} \vee \sigma(X_i^{(i,m)} \ : \ 1 \le m \le M_i), \\ \mathcal{F}^{(M)} & := & \mathcal{F}^{(*)} \vee \sigma(X_j^{(i,m)}, \Delta W_j^{(i,m)} \ : \ 0 \le i, j \le N \ ; \ 1 \le m \le M_i). \end{array}\right\} \tag{3.4.5}$$

Definition 3.4.7. *For every $i \in \{0, \ldots, N-1\}$, let $\mathbb{E}_i^*[\cdot]$ (resp. $\mathbb{P}_i^*$) be the conditional expectation (resp. conditional probability) with respect to $\mathcal{F}_i^{(*)}$, $\mathbb{E}_i^M[\cdot]$ (resp. $\mathbb{P}_i^M$) with respect to $\mathcal{F}_i^{(M)}$, and $\mathbb{E}^M[\cdot]$ (resp. $\mathbb{P}^M$) with respect to $\mathcal{F}^{(M)}$.*

Regression coefficients and approximate solutions. Recall the multilevel approximations $\big(y^{(\kappa,M)}(x), z^{(\kappa,M)}(x)\big)$ of Section 3.2 and, for $i \in \{1, \ldots, N\}$, define the functions

$$\bar{f}_i^{(M)}(x, y, z) := f(t_i, x, y_i^{(\kappa,M)}(x) + y, z_i^{(\kappa,M)}(x) + z)\mathbf{1}_{i<N} + f(T, x, \Phi(x) + y, 0)\mathbf{1}_{i=N}. \tag{3.4.6}$$

Note that $\bar{f}_i^{(M)}(\cdot)$ is like $\bar{f}_i(\cdot)$, except that $y_i^{(\kappa)}(x)$ (resp. $z_i^{(\kappa)}(x)$) has been replaced by $y_i^{(\kappa,M)}(x)$ (resp. $z_i^{(\kappa,M)}(x)$). For $i = 0, \ldots, N-1$ and $l = 1, \ldots, q$, define

$$\left.\begin{array}{rcl} \Psi_i^{(m)} & := & \sum_{j=i+1}^N \bar{f}_j^{(M)}(X_j^{(i,m)}, \bar{y}_j^{(M)}(X_j^{(i,m)}), \bar{z}_j^{(M)}(X_j^{(i,m)}))\Delta, \\ \Xi_{l,i}^{(m)} & := & \Delta W_{l,i}^{(m)} \sum_{j=i+1}^N \bar{f}_j^{(M)}(X_j^{(i,m)}, \bar{y}_j^{(M)}(X_j^{(i,m)}), \bar{z}_j^{(M)}(X_j^{(i,m)})). \end{array}\right\} \tag{3.4.7}$$

Recall the definition of $LS(\cdot)$ in (3.3.4). Let $\alpha_{0,i}^{(M)}$ be the $\mathbb{R}^{K_{0,i}^{(M)}}$-valued random variable solving $LS(X_i^{(i,1:M_i)}, \Psi_i^{(1:M_i)}, p_{0,i}(x))$ and $\alpha_{l,i}^{(M)}$ be the $\mathbb{R}^{K_{l,i}^{(M)}}$-valued random variable the solving $LS\big(X_i^{(i,1:M_i)},$
$\Xi_{l,i}^{(1:M_i)}, p_{l,i}(x)\big)$.

Definition 3.4.8. *Using the almost sure bounds of Proposition 3.4.6, define, for each $i \in \{0, \dots, N-1\}$, the approximation functions by*

$$\left.\begin{array}{rcl} \bar{y}_i^{(M)}(x) & := & -C(T-t_i) \vee \alpha_{0,i}^{(M)} \cdot p_{0,i}(x) \wedge C(T-t_i), \\ \\ \bar{z}_{l,i}^{(M)}(x) & := & -\frac{C(T-t_i)}{\Delta^{1/2}} \vee \alpha_{l,x}^{(M)} \cdot p_{l,i}(x) \wedge \frac{C(T-t_i)}{\Delta^{1/2}}, \\ \\ \bar{z}_i^{(M)}(x) & := & \big(\bar{z}_{1,i}^{(M)}(x), \dots, \bar{z}_{q,i}^{(M)}(x)\big). \end{array}\right\} \tag{3.4.8}$$

Corollary 3.4.9. *For all $i \in \{1, \dots, N\}$, $|\bar{f}_i^{(M)}(x, \bar{y}_i^{(M)}(x), \bar{z}_i^{(M)}(x))|$ is bounded by C for all $x \in \mathbb{R}^d$.*

Proof. Since $|\bar{y}_i^{(M)}(\cdot)|$, $|\bar{z}_i^{(M)}(\cdot)|$, $|y_i^{(\kappa,M)}(\cdot)|$ and $|z_i^{(\kappa,M)}(\cdot)|$ are bounded by C for all $i \in \{0, \dots, N\}$, the uniform Lipschitz continuity and bounds of $f(t_j, x, y, z)$ given in $(\mathbf{A_f})$ imply that $|\bar{f}_i^{(M)}(x, \bar{y}_i^{(M)}(x),$
$\bar{z}_i^{(M)}(x))|$ is also bounded by C for all i. $\square$

3.4.4 Error analysis

As in Section 3.3.2, the purpose of this section is to find a converging upper bound for the error term

$$\max_{0 \le i \le N-1} \mathbb{E}[\frac{1}{M_i} \sum_{m=1}^{M_i} |\bar{y}_i(X_i^{(i,m)}) - \bar{y}_i^{(M)}(X_i^{(i,m)})|^2] + \sum_{i=0}^{N} \mathbb{E}[\frac{1}{M_i} \sum_{m=1}^{M_i} |\bar{z}_i(X_i^{(i,m)}) - \bar{z}_i^{(M)}(X_i^{(i,m)})|^2]\Delta. \tag{3.4.9}$$

We introduce the following *random norms* to reduce notational complexity:

Definition 3.4.10. *For an $\mathcal{F}^{(NL)} \otimes \mathcal{B}(\mathbb{R}^d)$-measurable function $f : \mathbb{R}^d \times \Omega \to \mathbb{R}$, define the random norms*

$$\|f\|_{i,M}^2 := \frac{1}{M_i} \sum_{m=1}^{M_i} |f(X_i^{(i,m)})|^2 \quad \textit{and} \quad \|f\|_{i,\infty}^2 := \int |f(x)|^2 \mathbb{P} \circ X_i^{-1}(dx).$$

If the function f in Definition 3.4.10 is $\mathcal{F}^M \otimes \mathcal{B}(\mathcal{R}^d)$-measurable, then $\|f\|_{i,\infty}^2 = \mathbb{E}^M[f(X_i)]$ follows from Lemma 3.6.1.

Although the error in (3.4.9) is in terms of the random norms, we can relate it to the usual norm.

Definition 3.4.11. *For every $i \in \{0, \dots, N-1\}$ and $(\varepsilon_Y, \varepsilon_Z) \in (0, \infty)^2$, let*

$$\begin{aligned} B_{Y,i}^{(M)} &:= \{\ \varepsilon_Y \le \|\bar{y}_i(\cdot) - \bar{y}_i^{(M)}(\cdot)\|_{i,\infty}^2 - \|\bar{y}_i(\cdot) - \bar{y}_i^{(M)}(\cdot)\|_{i,M}^2\}, \\ B_{Z,i}^{(M)} &:= \{\exists l \in \{1, \dots, q\} : \varepsilon_Z \le \|\bar{z}_{l,i}(\cdot) - \bar{z}_{l,i}^{(M)}(\cdot)\|_{i,\infty}^2 - \|\bar{z}_{l,i}(\cdot) - \bar{z}_{l,i}^{(M)}(\cdot)\|_{i,M}^2\}. \end{aligned}$$

Due to Proposition 3.4.6, the random functions $|\bar{y}_i(X_i)(x) - \bar{y}_i^{(M)}(x)|$ and $|\bar{z}_{l,i}(X_{l,i})(x) - \bar{z}_{l,i}^{(M)}(x)|$ are bounded, therefore

$$\mathbb{E}[|\bar{y}_i(X_i) - \bar{y}_i^{(M)}(X_i)|^2] \leq \varepsilon_Y + \mathbb{E}[\|\bar{y}_i(\cdot) - \bar{y}_i^{(M)}(\cdot)\|_{i,M}^2] + C\mathbb{P}(B_{Y,i}^{(M)}),$$
$$\mathbb{E}[|\bar{z}_i(X_i) - \bar{z}_i^{(M)}(X_i)|^2] \leq \varepsilon_Z + \mathbb{E}[\|\bar{z}_i(\cdot) - \bar{z}_i^{(M)}(\cdot)\|_{i,M}^2] + C\mathbb{P}(B_{Z,i}^{(M)})$$

The following concentration of measure inequalities can be used to control the probabilities in the above inequalities.

Lemma 3.4.12. *There exists a constant C such that, if $\varepsilon_Y + \varepsilon_Z \leq C$, then for all $i \in \{0, \ldots, N-1\}$*

$$\mathbb{P}(B_{Y,i}^{(M)}) \leq C \exp\big(CK_{0,k}^{(M)} \log\big(\frac{(T-t_i)^2}{\varepsilon_Y}\big) - C\frac{M_i \varepsilon_Y}{(T-t_i)^2}\big),$$
$$\mathbb{P}(B_{Z,i}^{(M)}) \leq C \sum_{l=1}^{q} \exp\big(CK_{l,k}^{(M)} \log\big(\frac{(T-t_i)^2}{\Delta\varepsilon_Z}\big) - C\frac{M_i \Delta\varepsilon_Z}{(T-t_i)^2}\big).$$

Proof. Due to Proposition 3.4.6, the random functions $|\bar{y}_i(X_i)(x) - \bar{y}_i^{(M)}(x)|$ (resp. $|\bar{z}_{l,i}(X_{l,i})(x) - \bar{z}_{l,i}^{(M)}(x)|$) are bounded in i and x uniformly bounded by $C(T-t_i)$ (resp. C). The bounds are then obtained by the same technique as Theorem 2.4.5 of Chapter 2 for the equivalent sets ($C_k^{Y,M}$ and $C_k^{Z,M}$ in the notation of that article); one must only update the bounds of the random functions in that proof. □

Therefore, the usual error term

$$\max_{0 \leq i \leq N-1} \mathbb{E}[|\bar{y}_i(X_i i) - \bar{y}_i^{(M)}(X_i)|^2] + \sum_{i=0}^{N-1} \mathbb{E}[|\bar{z}_i(X_i) - \bar{z}_i^{(M)}(X_i)|^2]\Delta.$$

is controlled by controlling the error term (3.4.9).

Before presenting the main results of this section, the following notation will be useful.

Intermediate processes. As in the proof of Theorem 2.4.4 of Chapter 2, the error analysis of (3.4.9) makes use of intermediate processes.

Recall the definition (3.3.4) of $LS(\cdot)$. For $i \in \{0, \ldots, N-1\}$ and $l \in \{1, \ldots, q\}$, define

$$\bar{\Psi}_i^{(m)} := \sum_{j=i+1}^{N} \bar{f}_j(X_j^{(i,m)}, \bar{y}_j(X_j^{(i,m)}), \bar{z}_j(X_j^{(i,m)}))\Delta,$$
$$\bar{\Xi}_{l,i}^{(m)} := \Delta W_{l,i}^{(m)} \sum_{j=i+1}^{N} \bar{f}_j(X_j^{(i,m)}, \bar{y}_j(X_j^{(i,m)}), \bar{z}_j(X_j^{(i,m)})),$$

and let $\beta_{0,i}^{(M)}$ be the $\mathbb{R}^{K_{0,i}^{(M)}}$-valued random variable solving $LS\big(X_i^{(i,1:M_i)}, \bar{\Psi}_i^{(1:M_i)}, p_{0,i}(x)\big)$ and $\beta_{l,i}^{(M)}$ be the $\mathbb{R}^{K_{l,i}^{(M)}}$-valued random variable solving $LS\big(X_i^{(i,1:M_i)}, \bar{\Xi}_{l,i}^{(1:M_i)}, p_{l,i}(x)\big)$. The coefficients $\beta_{l,i}^{(M)}$ differ from $\alpha_{l,i}^{(M)}$ in that $\bar{y}_j(x)$ (resp. $\bar{z}_j(x)$) replaces $\bar{y}_j^{(M)}(x)$ (resp. $\bar{z}_j^{(M)}(x)$), and $\bar{f}_j$ replaces $\bar{f}_j^{(M)}$. Notice that the coefficients $\beta_{l,i}^{(M)}$ are not $\mathcal{F}_i^M$-measurable, because they depend on the data $\{X_j^{(i,m)}; j > i\}$.

Lemma 3.4.13. *For $i \in \{0, \ldots, N-1\}$ and $l \in \{1, \ldots, q\}$, $\mathbb{E}_i^M[\beta_{0,i}^{(M)}]$ solves $LS\big(X_i^{(i,1:M_i)}, \bar{y}_i(X_i^{(i,1:M_i)}), p_{0,i}(x)\big)$ and $\mathbb{E}_i^M[\beta_{l,i}^{(M)}]$ solves $LS\big(X_i^{(i,1:M_i)}, \bar{z}_{l,i}(X_i^{(i,1:M_i)}), p_{l,i}(x)\big)$.*

Proof. Using Lemma 3.6.1 in the same way as in Lemma 3.2.5, it follows that $\mathbb{E}_i^M[\bar{\Psi}_i^{(m)}] = \bar{y}_i(X_i^{(i,m)})$ and $\mathbb{E}_i^M[\bar{\Xi}_i^{(m)}] = \bar{z}_i(X_i^{(i,m)})$. The result then follows from Proposition 3.6.4(iii). $\square$

Recall the definition of the Markov chain $X^{(i,x)}$ in Section 3.1.3. For $i \in \{0, \ldots, N-1\}$, let $\mu_i^x := \mathbb{P} \circ (X_{i+1}^{(i,x)}, \ldots, X_N^{(i,x)})^{-1}$ and $\lambda_i^x := \mathbb{P} \circ (\Delta W_i, X_{i+1}^{(i,x)}, \ldots, X_N^{(i,x)})^{-1}$, and define the random functions

$$\left.\begin{array}{rcl} \mathcal{Y}_i(x) & := & \Delta \sum_{j=i+1}^N \int \bar{f}_j^{(M)}(\bar{y}_j^{(M)}(x_j), \bar{z}_j^{(M)}(x_j)) d\mu_i^x, \\ \mathcal{Z}_i(x) & := & \sum_{j=i+1}^N \int w \bar{f}_j^{(M)}(\bar{y}_j^{(M)}(x_j), \bar{z}_j^{(M)}(x_j)) d\lambda_i^x. \end{array}\right\} \tag{3.4.10}$$

Notice that the coefficients $\alpha_{l,i}^{(M)}$ are not $\mathcal{F}_i^M$-measurable, because they depend on the data $\{X_j^{(i,m)}; j > i\}$.

Lemma 3.4.14. *For all $i \in \{0, \ldots, N-1\}$ and $l \in \{1, \ldots, q\}$, $\mathbb{E}_i^M[\alpha_{0,l}^{(M)}]$ solves $LS\big(X_i^{(i,1:M_i)}, \mathcal{Y}_i(X_i^{(i,1:M_i)}), p_{0,i}(x)\big)$ and $\mathbb{E}_i^M[\alpha_{l,i}^{(M)}]$ solves $LS\big(X_i^{(i,1:M_i)}, \mathcal{Z}_{l,i}(X_i^{(i,1:M_i)}), p_{0,i}(x)\big)$, where $\mathcal{Z}_{l,i}(x)$ is the l-th component of $\mathcal{Z}_i(x)$.*

Moreover, the processes $\big(\mathcal{Y}_i(X_i), \mathcal{Z}_i(X_i)\big)_{0 \le i \le N}$ solve a discrete BSDE with driver $f_j(y,z) = \bar{f}_j^{(M)}(X_j, \bar{y}_j^{(M)}(X_j), \bar{z}_j^{(M)}(X_j))$ and terminal condition 0 under the filtration $(\mathcal{F}^M \vee \mathcal{F}_i^{(W)})_{i \ge 0}$.

Proof. Recall the definition of $\Psi_i^{(m)}$ and $\Xi_{l,i}^{(m)}$ in (3.4.7), and the σ-algebras in (3.4.5). The random functions $(\bar{f}_j^{(M)}(x,y,z), \bar{y}_j^{(M)}(x), \bar{z}_j^{(M)}(x))_{i+1 \le j \le N}$ are $\mathcal{F}_i^{(*)}$-measurable. Using Lemma 3.6.1 in the same way as in Lemma 3.2.5, it follows that

$$\begin{aligned} \mathcal{Y}_i(X_i^{(m)}) &= \sum_{j=i+1}^N \mathbb{E}_i^M[\bar{f}_j^{(M)}(\bar{y}_j^{(M)}(X_j^{(m)}), \bar{z}_j^{(M)}(X_j^{(m)}))]\Delta &&= \mathbb{E}_i^M[\Psi_i^{(m)}], \\ \mathcal{Z}_{l,i}(X_i^{(m)}) &= \sum_{j=i+1}^N \mathbb{E}_i^M[\Delta W_{l,i}^{(m)} \bar{f}_j^{(M)}(\bar{y}_j^{(M)}(X_j^{(m)}), \bar{z}_j^{(M)}(X_j^{(m)}))] &&= \mathbb{E}_i^M[\Xi_{l,i}^{(m)}]. \end{aligned}$$

The first result follows by Proposition 3.6.4(iii).

The second result is obtained in the same way as the first equality above:

$$\begin{aligned} \mathcal{Y}_i(X_i) &= \sum_{j=i+1}^N \mathbb{E}[\bar{f}_j^{(M)}(\bar{y}_j^{(M)}(X_j^{(m)}), \bar{z}_j^{(M)}(X_j^{(m)}))|\mathcal{F}^M \vee \mathcal{F}_i^{(W)}]\Delta, \\ \mathcal{Z}_{l,i}(X_i^{(m)}) &= \sum_{j=i+1}^N \mathbb{E}[\Delta W_{l,i}^{(m)} \bar{f}_j^{(M)}(\bar{y}_j^{(M)}(X_j^{(m)}), \bar{z}_j^{(M)}(X_j^{(m)}))|\mathcal{F}^M \vee \mathcal{F}_i^{(W)}]. \end{aligned}$$

$\square$

Lemma 3.4.15. *For all $i \in \{0, \ldots, N-1\}$ and $x \in \mathbb{R}^d$,*

$$|\mathcal{Y}_i(x)| \le C(T - t_i), \quad |\mathcal{Z}_i(x)| \le \frac{C(T-t_i)}{\Delta^{1/2}} \quad \forall x \in \mathbb{R}^d. \tag{3.4.11}$$

Proof. It was shown in Corollary 3.4.9 that $|\bar{f}_i^{(M)}(x, \bar{y}_i^{(M)}(x), \bar{z}_i^{(M)}(x))| \le C$ for all i. Therefore, Minkowski's inequlity implies that

$$
\begin{aligned}
|\mathcal{Y}_i(x)| &\le \Delta \sum_{j=i+1}^{N} \int |\bar{f}_j^{(M)}(\bar{y}_j^{(M)}(x_j), \bar{z}_j^{(M)}(x_j))| d\mu_i^x \le C(T - t_i), \\
|\mathcal{Z}_i(x)| &\le \sum_{j=i+1}^{N} \int |w| |\bar{f}_j^{(M)}(\bar{y}_j^{(M)}(x_j), \bar{z}_j^{(M)}(x_j))| d\lambda_i^x \le \frac{C(T - t_i)\big(\mathbb{E}[(\Delta W_i)^2]\big)^{1/2}}{\Delta}
\end{aligned}
$$

as required. $\square$

The following events will measure large deviations of $\big| \|\bar{y}_i(\cdot) - \mathcal{Y}_i(\cdot)\|_{i,\infty}^2 - \|\bar{y}_i(\cdot) - \mathcal{Y}_i(\cdot)\|_{i,M}^2 \big|$.

Definition 3.4.16. *For every $i \in \{0, \dots, N-1\}$ and $(\varepsilon_Y, \varepsilon_Z) \in (0, \infty)^2$, define*

$$
\begin{aligned}
C_{Y,i}^{(M)} &:= \{\, \varepsilon_Y \le \big| \|\bar{y}_i(\cdot) - \mathcal{Y}_i(\cdot)\|_{i,\infty}^2 - \|\bar{y}_i(\cdot) - \mathcal{Y}_i(\cdot)\|_{i,M}^2 \big| \}, \\
C_{Z,i}^{(M)} &:= \{\exists l \in \{1, \dots, q\} : \varepsilon_Z \le \big| \|\bar{z}_{l,i}(\cdot) - \mathcal{Z}_{l,i}(\cdot)\|_{i,\infty}^2 - \|\bar{z}_{l,i}(\cdot) - \mathcal{Z}_{l,i}(\cdot)\|_{i,M}^2 \big| \}.
\end{aligned}
$$

Recall the σ-algebras (3.4.5) and the conditional probabilities in Definition 3.4.7. The probability $\mathbb{P}_i^*$ of the events in Definition 3.4.16 is controlled by concentration of measure inequalities given in the following lemma:

Lemma 3.4.17. *There is a constant C such that for every $i \in \{0, \dots, N-1\}$ and $(\varepsilon_Y, \varepsilon_Z) \in (0, \infty)^2$,*

$$
\mathbb{P}_i^*(C_{Y,i}^{(M)}) \le 2 \exp\Big(- \frac{C \varepsilon_Y^2 M_i}{(T - t_i)^2} \Big), \quad \mathbb{P}_i^*(C_{Z,i}^{(M)}) \le 2q \exp\Big(- \frac{C \Delta \varepsilon_Z^2 M_i}{(T - t_i)^2} \Big). \tag{3.4.12}
$$

Proof. In Lemma 3.4.15 it was shown that the random functions $|\bar{y}_i(x) - \mathcal{Y}_i(x)|$ (resp. $|\bar{z}_{l,i}(x) - \mathcal{Z}_{l,i}(x)|$) are bounded by $C(T - t_i)$ (resp. $C(T - t_i)/\Delta^{1/2}$) uniformly in x. Moreover, for $i = 0, \dots, N$, the functions $\omega \mapsto \mathcal{Y}_i(\omega, x), \mathcal{Z}_i(\omega, x)$ are $\mathcal{F}_i^*$-measurable, and so $\{\bar{y}_i(X_i^{(i,m)}) - \mathcal{Y}_i(X_i^{(i,m)}) : 1 \le m \le M_i\}$ and $\{\bar{z}_i(X_i^{(i,m)}) - \mathcal{Z}_i(X_i^{(i,m)}) : 1 \le m \le M_i\}$ are $\mathcal{F}_i^M$-conditionally independent in the sense of Lemma 3.6.3. This means we can apply the conditional Hoeffding inequality, Lemma 3.6.3, to complete the proof. $\square$

Bias and variance. In order to express the error (3.4.9) in terms of the basis and the number of simulations, we make use error terms that are related to the bias due to the basis selection and the variance of each least-squares regression, as in Definition 3.3.6.

Definition 3.4.18. *For each $i \in \{0, \ldots, N-1\}$, define*

$$T_{1,i}^{(Y,M)} := \mathbb{E}_i^*[\inf_{\alpha \in \mathbb{R}^{K_{0,i}^{(M)}}} \|\alpha \cdot p_{0,i} - \bar{y}_i(\cdot)\|_{i,M}^2], \tag{3.4.13}$$

$$T_{1,i}^{(Z,M)} = \sum_{l=1}^{q} \mathbb{E}_i^*[\inf_{\alpha \in \mathbb{R}^{K_{l,i}^{(M)}}} \|\alpha \cdot p_{l,i} - \bar{z}_{l,i}(\cdot)\|_{i,M}^2], \tag{3.4.14}$$

$$T_{2,i}^{(Y,M)} = \mathbb{E}_i^M[\|(\alpha_{0,i}^{(M)} - \mathbb{E}_i^M[\alpha_{0,i}^{(M)}]) \cdot p_{0,i}\|_{i,M}^2], \tag{3.4.15}$$

$$T_{2,i}^{(Z,M)} = \sum_{l=1}^{q} \mathbb{E}_i^M[\|(\alpha_{l,i}^{(M)} - \mathbb{E}_i^M[\alpha_{l,i}^{(M)}]) \cdot p_{l,i}\|_{i,M}^2]. \tag{3.4.16}$$

Proposition 3.4.19. *For all $i \in \{0, \ldots, N-1\}$,*

$$T_{1,i}^{(Y,M)} \leq \inf_{\alpha \in \mathbb{R}^{K_{0,i}^{(M)}}} \mathbb{E}[|\alpha \cdot p_{0,i}(X_i) - \bar{y}_i(X_i)|^2], \tag{3.4.17}$$

$$T_{1,i}^{(Z,M)} \leq \sum_{l=1}^{q} \inf_{\alpha \in \mathbb{R}^{K_{l,i}^{(M)}}} \mathbb{E}[|\alpha \cdot p_{l,i}(X_i) - \bar{z}_{l,i}(X_i)|^2], \tag{3.4.18}$$

$$T_{2,i}^{(Y,M)} \leq \frac{C(T-t_i)^2 K_{0,i}^{(M)}}{M_i}, \quad T_{2,i}^{(Z,M)} \leq \frac{C(T-t_i)^2 \sum_{l=1}^{q} K_{l,i}^{(M)}}{\Delta M_i}. \tag{3.4.19}$$

Proof. The proofs of (3.4.17) and (3.4.18) are standard: see [LGW06, Proposition 4].

For $l \in \{0, \ldots, q\}$, let $P_{l,i}^{(M)}$ be the $M_i \times K_{l,i}^{(M)}$-dimensional random matrix with m-th row $p_{l,i}(X_i^{(i,m)})$ and recall the definitions of $\Psi_i^{(m)}$ and $\Xi_{l,i}^{(m)}$ in (3.4.7). Assume (without loss of generality) that $(P_{l,i}^{(M)})^\top P_{l,i}^{(M)}/M_i$ is the identity matrix. Using the method of Proposition 3.3.7, for each $m \in \{1, \ldots, M_i\}$, one can show that

$$\begin{aligned}
&\mathbb{E}_i^M[|p_{l,i}(X_i^{(k,m)}) \cdot (\alpha_{l,i}^{(M)} - \mathbb{E}_i^M[\alpha_{l,i}^{(M)}])|^2] \\
&= (p_{l,i}(X_i^{(i,m)}))^\top \frac{(P_{l,i}^{(M)})^\top}{M_i} \mathbb{E}_i^M[(V - \mathbb{E}_i^M V)(V - \mathbb{E}_i^M V)^\top] \frac{P_{l,i}^{(M)}}{M_i} p_{l,i}(X_i^{(i,m)})
\end{aligned}$$

where V is the M_i-dimensional random vector with m-th component $\Psi_i^{(m)} \mathbf{1}_{l=0} + \Xi_{l,i}^{(m)} \mathbf{1}_{l>0}$. The off-diagonal terms of the matrix $\mathbb{E}_i^M[(V - \mathbb{E}_i^M V)(V - \mathbb{E}_i^M V)^\top]$ are zero due to the independence of the samples, and the diagonal terms are dominated by $\mathbb{E}_i^M[|\Psi_i^{(m)}|^2]\mathbf{1}_{l=0} + \mathbb{E}_i^M[|\Xi_{l,i}^{(m)}|^2]\mathbf{1}_{l>0}$. Since $|\bar{f}_i^{(M)}(x, \bar{y}_i^{(M)}(x), \bar{z}_i^{(M)}(x))| \leq C$ for all i, it follows that

$$\begin{aligned}
\mathbb{E}_i^M[|\Psi_i^{(m)}|^2] &= \mathbb{E}_i^M[|\sum_{j=i+1}^{N} \bar{f}_j^{(M)}(X_j^{(i,m)}, \bar{y}_j^{(M)}(X_j^{(i,m)}), \bar{z}_j^{(M)}(X_j^{(i,m)}))\Delta|^2] \leq C(T-t_i)^2, \\
\mathbb{E}_i^M[|\Xi_{l,i}^{(m)}|^2] &= \mathbb{E}_i^M[|\Delta W_{l,i}^{(m)} \sum_{j=i+1}^{N} \bar{f}_j^{(M)}(X_j^{(i,m)}, \bar{y}_j^{(M)}(X_j^{(i,m)}), \bar{z}_j^{(M)}(X_j^{(i,m)}))|^2] \\
&\leq \frac{C(T-t_i)^2 \mathbb{E}[|\Delta W_i|^2]}{\Delta^2} \leq \frac{C(T-t_i)^2}{\Delta}
\end{aligned}$$

Since $\mathbb{E}_i^M[|\Psi_i^{(m)}|^2]\mathbf{1}_{l=0} + \mathbb{E}_i^M[|\Xi_{l,i}^{(m)}|^2]\mathbf{1}_{l>0}$ have upper bounds that are independent of m, one can complete the proof in the same way as the proof of Proposition 3.3.7. □

We are concerned with finding a determinisitc upper bound in terms of N, M, ε_Y and ε_Z for the error terms

$$\max_{0\le i\le N-1} \mathbb{E}[\|\bar{y}_i(\cdot) - \bar{y}_i^{(M)}(\cdot)\|_{i,M}^2] \quad \text{and} \quad \sum_{i=0}^{N-1} \mathbb{E}[\|\bar{z}_i(\cdot) - \bar{z}_i^{(M)}(\cdot)\|_{i,M}^2]\Delta.$$

Theorem 3.4.20. *Suppose that the multilevel approximation $(y^{(\kappa,M)}, z^{(\kappa,M)})$ is computed as in Section 3.3.3. For every $j \in \{0,\dots,N-1\}$, define*

$$\begin{aligned}
\mathcal{E}(Y,j) &:= \inf_{\alpha\in\mathbb{R}^{K_{0,j}^{(M)}}} \mathbb{E}[|\alpha\cdot p_{0,j}(X_j) - \bar{y}_j(X_j)|^2] + \frac{(T-t_i)^2K_{0,i}^{(M)}}{M_i} + \exp\Big(-\frac{C\varepsilon_Y^2 M_i}{(T-t_i)^2}\Big),\\
\mathcal{E}(Z,j) &:= \sum_{l=1}^{q}\left(\inf_{\alpha\in\mathbb{R}^{K_{l,i}^{(M)}}} \mathbb{E}[|\alpha\cdot p_{l,i}(X_i) - \bar{z}_{l,i}(X_i)|^2] + \frac{(T-t_i)^2K_{l,i}^{(M)}}{\Delta M_i} + \frac{1}{\Delta}\exp\Big(-\frac{C\Delta\varepsilon_Z^2 M_i}{(T-t_i)^2}\Big)\right)\\
\mathcal{E}(B,j) &:= \exp\big(CK_{0,k}^{(M)}\log\big(\frac{(T-t_i)^2}{\varepsilon_Y}\big) - C\frac{M_i\varepsilon_Y}{(T-t_i)^2}\big) + \sum_{l=1}^{q}\exp\big(CK_{l,i}^{(M)}\log\big(\frac{(T-t_i)^2}{\Delta\varepsilon_Z}\big) - C\frac{M_i\Delta\varepsilon_Z}{(T-t_i)^2}\big).
\end{aligned}$$

Then, for Δ sufficiently small, there exists a (possibly different) constant C such that, if $\varepsilon_Y+\varepsilon_Z \le C$, then for all $i\in\{0,\dots,N-1\}$

$$\mathbb{E}[\|\bar{y}_i(\cdot) - \bar{y}_i^{(M)}(\cdot)\|_{i,M}^2] \le CN^{-1} + C\varepsilon_Y + C\varepsilon_Z + \mathcal{E}(Y,i) + C\sum_{j=i+1}^{N}\{\mathcal{E}(Y,j)+\mathcal{E}(Z,j)+\mathcal{E}(B,j)\}\Delta, \tag{3.4.20}$$

$$\sum_{j=0}^{N}\mathbb{E}[\|\bar{z}_j(\cdot) - \bar{z}_j^{(M)}(\cdot)\|_{j,M}^2]\Delta \le CN^{-1} + C\varepsilon_Y + C\varepsilon_Z + C\sum_{j=0}^{N}\{\mathcal{E}(Y,j)+\mathcal{E}(Z,j)+\mathcal{E}(B,j)\}\Delta. \tag{3.4.21}$$

Proof. Using $\|\bar{y}_i(\cdot) - \bar{y}_i^{(M)}(\cdot)\|_{i,M}^2 \le \|\bar{y}_i(\cdot) - \alpha_{0,i}^{(M)}\cdot p_{0,i}\|_{i,M}^2$, Pythagorus' Theorem yields

$$\begin{aligned}
\|\bar{y}_i(\cdot) - \bar{y}_i^{(M)}(\cdot)\|_{i,M}^2 &\le \|\bar{y}_i(\cdot) - \mathbb{E}_i^M[\beta_{0,i}^{(M)}]\cdot p_{0,i}\|_{i,M}^2 + \|(\mathbb{E}_i^M[\beta_{0,i}^{(M)}] - \alpha_{0,i}^{(M)})\cdot p_{0,i}\|_{i,M}^2\\
&\le \|\bar{y}_i(\cdot) - \mathbb{E}_i^M[\beta_{0,i}^{(M)}]\cdot p_{0,i}\|_{i,M}^2 + 2\|\mathbb{E}_i^M[\beta_{0,i}^{(M)} - \alpha_{0,i}^{(M)}]\cdot p_{0,i}\|_{i,M}^2\\
&\quad + 2\|(\mathbb{E}_i^M[\alpha_{0,i}^{(M)}] - \alpha_{0,i}^{(M)})\cdot p_{0,i}\|_{i,M}^2
\end{aligned} \tag{3.4.22}$$

It is shown in Lemma 3.4.14 that $\mathbb{E}_i^M[\alpha_{0,i}^{(M)}]$ solves $LS\big(X_i^{(i,1:M_i)}, \mathcal{Y}_i(X_i^{(i,1:M_i)}), p_{0,i}(x)\big)$ and in Lemma 3.4.13 that $\mathbb{E}_i^M[\beta_{0,i}^{(M)}]$ solves $LS\big(X_i^{(i,1:M_i)}, \bar{y}_i(X_i^{(i,1:M_i)}), p_{0,i}(x)\big)$, therefore, by Proposition 3.6.4(i) that $\mathbb{E}_i^M[\alpha_{0,i}^{(M)} - \beta_{0,i}^{(M)}]$ solves $LS\big(X_i^{(i,1:M_i)}, (\bar{y}_i - \mathcal{Y}_i)(X_i^{(i,1:M_i)}), p_{0,i}(x)\big)$. It follows from Proposition 3.6.4(ii) that $\|\mathbb{E}_i^M[\beta_{0,i}^{(M)} - \alpha_{0,i}^{(M)}]\cdot p_{0,i}\|_{i,M}^2 \le \|\bar{y}_i(\cdot) - \mathcal{Y}_i(\cdot)\|_{i,M}^2$. Therefore, recalling the event $C_{Y,i}^{(M)}$ from Definition 3.4.16 and the bounds on $|\bar{y}_i(x) - \mathcal{Y}_i(x)|$ from Proposition 3.4.6 and Lemma 3.4.15,

$$\|\mathbb{E}_i^M[\beta_{0,i}^{(M)} - \alpha_{0,i}^{(M)}]\cdot p_{0,i}\|_{i,M}^2 \le \varepsilon_Y + \|\bar{y}_i(\cdot) - \mathcal{Y}_i(\cdot)\|_{i,\infty}^2 + C\mathbf{1}_{C_{Y,i}^{(M)}} \tag{3.4.23}$$

Since the function $\mathcal{Y}_i(x)$ are $\mathcal{F}^M\otimes\mathcal{B}(\mathbb{R}^d)$-measurable, it follows from Lemma 3.6.1 that $\|\bar{y}_i(\cdot) -$

$\mathcal{Y}_i(\cdot)\|_{i,\infty}^2 = \mathbb{E}[|\bar{y}_i(X_i) - \mathcal{Y}_i(X_i)|^2 \big| \mathcal{F}^M]$. Lemma 3.4.5 states that $\big(\bar{y}_i(X_i), \bar{z}_i(X_i)\big)_{0 \le i \le N}$ is a discrete BSDE with terminal condition 0 and driver $f_{1,j}(y,z) = \bar{f}_j(X_j, y, z)$, and Lemma 3.4.14 states that $\big(\mathcal{Y}_i(X_i), \mathcal{Z}_i(X_i)\big)$ is a discrete BSDE with terminal condition 0 and driver $f_{2,j}(y,z) = \bar{f}_j^{(M)}\big(X_j, \bar{y}_j^{(M)}(X_j), \bar{z}_j^{(M)}(X_j)\big)$; in both cases, we work with respect to the filtration $(\mathcal{F}^M \vee \mathcal{F}_i)_{i \ge 0}$. Set $\gamma = 12qL_f^2$ and define $\Gamma_i := (1+\gamma\Delta)^i$. It follows from Proposition 3.4.3 that $\|\bar{y}_i(\cdot) - \mathcal{Y}_i(\cdot)\|_{i,\infty}^2 \Gamma_i$ is bounded above by

$$\begin{aligned} &C \sum_{j=i+1}^{N} \Gamma_j (\frac{1}{\gamma} + \Delta) \mathbb{E}[|\bar{f}_j(X_j, \bar{y}_j(X_j), \bar{z}_j(X_j)) - \bar{f}_j^{(M)}(X_j, \bar{y}_j^{(M)}(X_j), \bar{z}_j^{(M)}(X_j))|^2 \big| \mathcal{F}^M]\Delta \\ &\le C \sum_{j=i+1}^{N} \Gamma_j (\frac{1}{\gamma} + \Delta) \mathbb{E}[|y_j^{(\kappa)}(X_j) - y_j^{(\kappa,M)}(X_j)|^2 + |z_j^{(\kappa)}(X_j) - z_j^{(\kappa,M)}(X_j)|^2 \big| \mathcal{F}^M]\Delta \\ &\quad + C \sum_{j=i+1}^{N} \Gamma_j (\frac{1}{\gamma} + \Delta) \mathbb{E}[|\bar{y}_j(X_j) - \bar{y}_j^{(M)}(X_j)|^2 + |\bar{z}_j(X_j) - \bar{z}_j^{(M)}(X_j)|^2 \big| \mathcal{F}^M]\Delta. \end{aligned}$$

Recalling the random norms $\|\cdot\|_{\kappa,i,\infty}$ from Definition 3.3.4, it follows that the right hand side of the the above inequality is equal to

$$\begin{aligned} &C \sum_{j=i+1}^{N} \Gamma_j (\frac{1}{\gamma} + \Delta) \{\|y_j^{(\kappa)}(\cdot) - y_j^{(\kappa,M)}(\cdot)\|_{\kappa,j,\infty}^2 + \|z_j^{(\kappa)}(\cdot) - z_j^{(\kappa,M)}(\cdot)\|_{\kappa,j,\infty}^2\}\Delta \\ &\quad + C \sum_{j=i+1}^{N} \Gamma_j (\frac{1}{\gamma} + \Delta) \{\|\bar{y}_j(\cdot) - \bar{y}_j^{(M)}(\cdot)\|_{j,\infty}^2 + \||\bar{z}_j(\cdot) - \bar{z}_j^{(M)}(\cdot)\|_{j,\infty}^2\}\Delta. \end{aligned}$$

Then, using the concentration of measure events from Definition 3.4.11 and (3.3.10) and the bounds on $(|y_j^{(\kappa)}(x)|, |z_j^{(\kappa)}(x)|)$ from Lemma 3.2.6 and the bounds on $(|\bar{y}_j(x)|, |\bar{z}_j(x)|)$ from Proposition 3.4.6, it follows that $\|\bar{y}_i(\cdot) - \mathcal{Y}_i(\cdot)\|_{i,\infty}^2 \Gamma_i$ is bounded above by

$$\begin{aligned} &C\varepsilon_Y + C\varepsilon_Z + C\varepsilon_Y^{(k)} + C\varepsilon_Z^{(k)} + C \sum_{j=i+1}^{N} \big\{\mathbf{1}_{B_{Y,j}^{(M)}} + \mathbf{1}_{B_{Z,j}^{(M)}} + \mathbf{1}_{A_{Y,j}^{(k)}} + \mathbf{1}_{A_{Z,j}^{(k)}}\big\}\Delta \\ &\quad + C \sum_{j=i+1}^{N} \Gamma_j (\frac{1}{\gamma} + \Delta) \{\|\bar{y}_j(\cdot) - \bar{y}_j^{(M)}(\cdot)\|_{j,M}^2 + \|\bar{z}_j(\cdot) - \bar{z}_j^{(M)}(\cdot)\|_{j,M}^2\}\Delta \\ &\quad + C \sum_{j=i+1}^{N} \{\|y_j^{(\kappa)}(\cdot) - y_j^{(\kappa,M)}(\cdot)\|_{\kappa,j,M}^2 + \|z_j^{(\kappa)}(\cdot) - z_j^{(\kappa,M)}(\cdot)\|_{\kappa,j,M}^2\}\Delta \end{aligned} \tag{3.4.24}$$

where we have used $\Gamma_j \le e^{\gamma T} = C$ in the first and last lines above. Observe that taking the expectation $\mathbb{E}$ in the sum in the last line is equal to $\mathcal{E}(N)$, which is defined in (3.3.30). Since the multilevel approximations $(y^{(\kappa,M)}, z^{(\kappa,M)})$ are computed as in Section 3.3.3, $\mathcal{E}(N) \le CN^{-1}$.

Now, multiplying by Γ_i in (3.4.22), taking the expectation $\mathbb{E}$, and substituting the results of

(3.4.23) and (3.4.24) implies that $\mathbb{E}[\|\bar{y}_i(\cdot) - \bar{y}_i^{(M)}(\cdot)\|_{i,M}^2 \Gamma_i]$ is bounded above by

$$C\varepsilon_Y + C\varepsilon_Z + CT_{1,i}^{(Y,M)} + CN^{-1} + CT_{2,i}^{(Y,M)} + C\mathbb{P}(C_{Y,i}^{(M)}) + C\sum_{j=i+1}^{N}\{\mathbb{P}(B_{Y,j}^{(M)}) + \mathbb{P}(B_{Z,j}^{(M)})\}\Delta$$
$$+ C\sum_{j=i+1}^{N}\Gamma_j(\frac{1}{\gamma} + \Delta)\mathbb{E}[\|\bar{y}_j(\cdot) - \bar{y}_j^{(M)}(\cdot)\|_{j,M}^2 + \|\bar{z}_j(\cdot) - \bar{z}_j^{(M)}(\cdot)\|_{j,M}^2]\Delta \tag{3.4.25}$$

Using the bounds on $T_{1,j}^{(Y,M)}$ and $T_{2,j}^{(Y,M)}$ from Proposition 3.4.19, the bounds on $\mathbb{P}(B_{Y,j}^{(M)})$ and $\mathbb{P}(B_{Z,j}^{(M)})$ from Lemma 3.4.17, and the bounds on $\mathbb{P}(B_{Y,j}^{(M)})$ and $\mathbb{P}(B_{Z,j}^{(M)})$ from Lemma 3.4.12, this implies that $\sum_{j=i}^{N}\Gamma_j\mathbb{E}[\|\bar{y}_j(\cdot) - \bar{y}_j^{(M)}(\cdot)\|_{j,M}^2]\Delta$ is bounded above by

$$CN^{-1} + C\varepsilon_Y + C\varepsilon_Z + C\sum_{j=i+1}^{N}\{\mathcal{E}(Y,j) + \mathcal{E}(B,j)\}\Delta$$
$$+ C\sum_{j=i}^{N}\Gamma_j(\frac{1}{\gamma} + \Delta)\mathbb{E}[\|\bar{y}_j(\cdot) - \bar{y}_j^{(M)}(\cdot)\|_{j,M}^2 + \|\bar{z}_j(\cdot) - \bar{z}_j^{(M)}(\cdot)\|_{j,M}^2]\Delta \tag{3.4.26}$$

Analogous computations yield that $\sum_{j=i}^{N}\Gamma_j\mathbb{E}[\|\bar{z}_j(\cdot) - \bar{z}_j^{(M)}(\cdot)\|_{j,M}^2]\Delta$ is bounded above by

$$CN^{-1} + C\varepsilon_Y + C\varepsilon_Z + C\sum_{j=i+1}^{N}\{\mathcal{E}(j) + \mathcal{E}(Z,j)\}\Delta$$
$$+ C\sum_{j=i}^{N}\Gamma_j(\frac{1}{\gamma} + \Delta)\mathbb{E}[\|\bar{y}_j(\cdot) - \bar{y}_j^{(M)}(\cdot)\|_{j,M}^2 + \|\bar{z}_j(\cdot) - \bar{z}_j^{(M)}(\cdot)\|_{j,M}^2]\Delta \tag{3.4.27}$$

Combining (3.4.26) and (3.4.27), it follows that

$$\sum_{j=i}^{N}\Gamma_j\mathbb{E}[\|\bar{y}_j(\cdot) - \bar{y}_j^{(M)}(\cdot)\|_{j,M}^2 + \|\bar{z}_j(\cdot) - \bar{z}_j^{(M)}(\cdot)\|_{j,M}^2]\Delta$$
$$\leq CN^{-1} + C\varepsilon_Y + C\varepsilon_Z + C\sum_{j=i+1}^{N}\{\mathcal{E}(Y,j) + \mathcal{E}(Z,j) + \mathcal{E}(B,j)\}\Delta$$
$$+ C\sum_{j=i}^{N}\Gamma_j(\frac{1}{\gamma} + \Delta)\mathbb{E}[\|\bar{y}_j(\cdot) - \bar{y}_j^{(M)}(\cdot)\|_{j,M}^2 + \|\bar{z}_j(\cdot) - \bar{z}_j^{(M)}(\cdot)\|_{j,M}^2]\Delta.$$

Selecting γ sufficiently large and Δ sufficiently small so that $(\frac{1}{\gamma} + \Delta) \leq \frac{1}{2C}$, it follows that

$$\sum_{j=i}^{N}\Gamma_j\mathbb{E}[\|\bar{y}_j(\cdot) - \bar{y}_j^{(M)}(\cdot)\|_{j,M}^2 + \|\bar{z}_j(\cdot) - \bar{z}_j^{(M)}(\cdot)\|_{j,M}^2]\Delta$$
$$\leq CN^{-1} + C\varepsilon_Y + C\varepsilon_Z + C\sum_{j=i+1}^{N}\{\mathcal{E}(Y,j) + \mathcal{E}(Z,j) + \mathcal{E}(B,j)\}\Delta.$$

Substituting $1 \leq \Gamma_i \leq C$ for all i into the inequality above completes the proof of (3.4.21), and substituting the inequality above into (3.4.24) completes the proof of (3.4.20). □

3.4.5 Complexity analysis

Much like Section 3.3.3, in this section we use the error analysis of Theorem 3.4.20 to calibrate the basis functions $\big\{p_{l,i}(x) : i \in \{0, \dots, N-1\}, l \in \{0, \dots, q\}\big\}$ and the number of simulations $\{M_0, \dots, M_{N-1}\}$ so that the error term (3.4.9) is bounded above by CN^{-1}. The inequalities (3.4.20) and (3.4.21) show that it is sufficient to choose these parameters so that $\varepsilon_Y + \varepsilon_Z + \mathcal{E}(Y,i) + \mathcal{E}(Z,i) + \mathcal{E}(B,i) \le CN^{-1}$. This implies that we set $\varepsilon_Y = \varepsilon_Z = CN^{-1}$.

Since there will be much repetition with Section 3.3.3, we will give only the most important details in this section and refer the reader to that section for a more detailed account.

Similarly Section 3.3.3, we use the inequalities of the form

$$\inf_{\alpha \in \mathbb{R}^{K^{(M)}_{0,i}}} \mathbb{E}[|\bar{y}_i(X_i) - \alpha \cdot p_{0,i}(X_i)|^2] \le CN^{-1} + C \inf_{\alpha \in \mathbb{R}^{K^{(M)}_{0,i}}} \mathbb{E}[|\bar{y}_{t_i}(X_{t_i}) - \alpha \cdot p_{0,i}(X_{t_i})|^2].$$

in order to take advantage of the smoothness properties of the functions $\bar{y}_{t_i}(\cdot)$ and $\bar{z}_{t_i}(\cdot)$ in the selection of the basis functions.

Proposition 3.4.21. *Let $n \in \{2, 3, \dots\}$ and assume $(\mathbf{A}'_{\mathbf{diff}})$. There is a choice of basis functions and simulation sizes $(M_0, \dots, M_N)$ and a constant C (independent of N) with dependencies as in $(\mathbf{A}_{\mathbf{diff}})$ so that, for every N sufficiently large, the error*

$$\max_{0 \le i \le N-1} \mathbb{E}[\|\bar{y}_i(\cdot) - \bar{y}^{(M)}_i(\cdot)\|^2_{i,M}] + \sum_{j=0}^{N} \mathbb{E}[\|\bar{z}_j(\cdot) - \bar{z}^{(M)}_j(\cdot)\|^2_{j,M}]\Delta$$

is bounded above by CN^{-1} for an algorithm complexity bounded above by

$$\mathcal{C} = CN^{1+\frac{d}{2}} \log^{d+1}(N).$$

Proof. We take $p_{0,i}$ to be local polynomials of degree n on hypercubes, $p_{0,i}$ to be local polynomials of degree $n-1$ on hypercubes, as in Section 3.3.3. Using the gradient bounds (3.4.2) instead of (3.2.6), we use the Taylor expansion to calibrate the hypercube lengths $\delta_{y,i} = CN^{-\frac{1}{2n}}(T-t_i)^{\frac{n-2}{2n}}$ and $\delta_{z,i} = N^{-\frac{1}{2(n-1)}}(T-t_i)^{\frac{n-2}{2(n-1)}}$. Therefore,

$$K^{(M)}_{0,i} = CN^{\frac{d(n-1)}{2n}}(N-i)^{-\frac{d(n-2)}{2n}} \log^d(N) \quad \text{and} \quad K^{(M)}_{l,i} = CN^{\frac{d}{2}}(N-i)^{-\frac{d(n-2)}{2(n-1)}} \log^d(N).$$

From $\mathcal{E}(Y,i) \le N^{-1}$, the number of simulations must dominate the number of simulations M_i must dominate $C(N-i)^2 K^{(M)}_{0,i}/N$, whereas, from $\mathcal{E}(Z,i) \le CN^{-1}$, M_i must dominate $\sum_{l=1}^{q} C(N-i)^2 K^{(M)}_{l,i}$. We have used $T - t_i = \frac{T}{N}(N-i)$. The later term is greater, so, substituting the value of $K^{(M)}_{l,i}$ computed above, M_i must dominate

$$CN^{\frac{d}{2}}(N-i)^{2-\frac{d(n-2)}{2(n-1)}} \log^d(N).$$

The exponential terms in $\mathcal{E}(Z,i)$ require that M_i dominates $CN(N-i)^2$, but this is smaller in high dimension $d > 2$. Finally, the exponential terms in $\mathcal{E}(B,i)$ also set a requirement on M_i, but this is the same as the requirement from $\mathcal{E}(Y,i)$ and $\mathcal{E}(Z,i)$, whence we take

$$M_i = CN^{\frac{d}{2}}(N-i)^{2-\frac{d(n-2)}{2(n-1)}}\log^d(N).$$

The numerical effort for this algorithm is dominated by simulating the data. For each $i \in \{0,\dots,N-1\}$, we must simulate M_i paths of X, which costs in total $CN\sum_{i=0}^{N-1} M_i$ flops. This is more than the total cost of the sorting algorithm, which is $C\log(N)\sum_{i=0}^{N-1} M_i$. Therefore, the complexity is dominated by

$$\mathcal{C} = CN^{1+\frac{d}{2}}(1+\int_0^{N-1}(N-t)^{2-\frac{d(n-2)}{2(n-1)}}dt)\log^d(N) \le C(N^{1+\frac{d}{2}} + N^{4+\frac{d}{2(n-1)}}\mathbf{1}_{d<6(n-1)/(n-2)})\log^{1+d}(N).$$

□

As in the Section 3.3.3, additional smoothness of the terminal condition may lead to efficiency improvements. The following additional assumptions mirror $(\mathbf{A}_{\partial\mathbf{\Phi}})$ and Lemma 3.2.3.

3.5 Approximating the BSDE without decomposition

In this section, we use the techniques of Section 3.4 to approximate the full BSDE (Y,Z) in (3.1.4) and compare it's efficiency to that of approximating of the system 3.1.1 combined with the multilevel scheme of Section 3.2 for the linear part. Since there is much overlap with Section 3.4, we we use the same notation and summarize only the main results.

First, note that $Y_t = y_t(X_t) + \bar{y}_t(X_t) =: Y_t(X_t)$ and $Z_t = z_t(X_t) + \bar{z}_t(X_t) =: Z_t(X_t)$. Under the assumption $(\mathbf{A}'_{\mathbf{diff}})$, it follows, using the gradient bounds (3.2.6) and (3.4.2), that the functions $Y_t(\cdot)$ are n-times continuously differentiable, and $Z_t(\cdot)$ are $(n-1)$-times continuously differentiable, and that the gradient satisfy the bounds

$$\left.\begin{array}{rcll} \left|\frac{\partial^r}{\partial x_{i_1}\dots\partial x_{i_r}}Y_t(x)\right| & \le & C(T-t)^{(1-r)/2}, & \text{for } r=1,\dots,n, \\ \left|\frac{\partial^r}{\partial x_{i_1}\dots\partial x_{i_r}}Z_t(x)\right| & \le & C(T-t)^{-r/2}, & \text{for } r=1,\dots,n-1 \end{array}\right\} \text{for all } t\in[0,T). \tag{3.5.1}$$

We use a discrete BSDE $(Y_i, Z_i)_{0\le i\le N}$ - see Definition 3.4.2 - with terminal condition $\Phi(X_N)$ and driver $f_j(y,z) = f(t_j, X_j, y, z)$ to discretize the continuous BSDE in (3.1.4). By analogue to Lemma 3.4.5, there are deterministic, measurable functions $Y_i(x)$ and $Z_i(x)$ such that $Y_i = Y_i(X_i)$ and $Z_i = Z_i(X_i)$ almost surely for all $i \in \{0\dots,N\}$. By analogue to Proposition 3.4.6, the bounds $|Y_i(x)| \le C$ and $|Z_i(x)| \le C$ hold for all $i \in \{0\dots,N\}$. We use Monte Carlo least-squares in the same way as in Section 3.4.3 to approximate the function $Y_i(x)$ (resp. $Z_i(x)$) with a $\mathcal{F}^M\otimes\mathcal{B}(\mathbb{R}^d)$-measurable function $Y_i^{(M)}(x)$ (resp. $Z_i^{(M)}(x)$). The following theorem describes the error due to this approximation, and is analogous to Theorem 3.4.20.

Theorem 3.5.1. *For every $j \in \{0, \dots, N-1\}$, define*

$$\begin{aligned}
\bar{\mathcal{E}}(Y,j) &:= \inf_{\alpha \in \mathbb{R}^{K_{0,j}^{(M)}}} \mathbb{E}[|\alpha \cdot p_{0,j}(X_j) - Y_j(X_j)|^2] + \frac{K_{0,i}^{(M)}}{M_i} + \exp\Big(- C\varepsilon_Y^2 M_i\Big),\\
\bar{\mathcal{E}}(Z,j) &:= \sum_{l=1}^{q} \left(\inf_{\alpha \in \mathbb{R}^{K_{l,i}^{(M)}}} \mathbb{E}[|\alpha \cdot p_{l,i}(X_i) - Z_{l,i}(X_i)|^2] + \frac{K_{l,i}^{(M)}}{\Delta M_i} + \frac{1}{\Delta} \exp\Big(- C\Delta\varepsilon_Z^2 M_i\Big)\right)\\
\bar{\mathcal{E}}(B,j) &:= \exp\big(- CK_{0,k}^{(M)} \log\big(\varepsilon_Y\big) - CM_i\varepsilon_Y\big) + \sum_{l=1}^{q} \exp\big(- CK_{l,k}^{(M)} \log(\varepsilon_Z) - CM_i\varepsilon_Z\big).
\end{aligned}$$

Then, for Δ sufficiently small, there exists a constant C such, if $\varepsilon_Y + \varepsilon_Z \le C$, then for all $i \in \{0, \dots, N-1\}$

$$\mathbb{E}[\|Y_i(\cdot) - Y_i^{(M)}(\cdot)\|_{i,M}^2] \le C\varepsilon_Y + C\varepsilon_Z + \bar{\mathcal{E}}(Y,i) + C\sum_{j=i+1}^{N} \{\bar{\mathcal{E}}(Y,j) + \bar{\mathcal{E}}(Z,j) + \bar{\mathcal{E}}(B,j)\}\Delta, \tag{3.5.2}$$

$$\sum_{j=0}^{N} \mathbb{E}[\|Z_j(\cdot) - Z_j^{(M)}(\cdot)\|_{j,M}^2]\Delta \le C\varepsilon_Y + C\varepsilon_Z + C\sum_{j=0}^{N}\{\bar{\mathcal{E}}(Y,j) + \bar{\mathcal{E}}(Z,j) + \bar{\mathcal{E}}(B,j)\}\Delta. \tag{3.5.3}$$

The main difference between the $\bar{\mathcal{E}}(\cdots)$ terms of the above theorem to the equivalent terms $\mathcal{E}(\cdots)$ in Theorem 3.4.20 is that the terms $(T - t_i)$ are gone. The reason for this difference is, roughly speaking, that the almost sure bounds of $\bar{y}_i$ are $C(T-t_i)$ and those of $\bar{z}_i$ are $C(T-t_i)^2/\Delta$, whereas the almost sure bounds of Y_i and Z_i are bounded uniformly by C in i. This makes a big impact on the complexity of the algorithm, as shown in the following proposition.

Proposition 3.5.2. *Let $n \in \{2, 3, \dots\}$ and assume $(\mathbf{A}'_{\mathbf{diff}})$. There is a choice of basis functions and simulation sizes $(M_0, \dots, M_N)$ and a constant C (independent of N) as in $(\mathbf{A}_{\mathbf{diff}})$ so that, for every N sufficiently large, the error*

$$\max_{0 \le i \le N-1} \mathbb{E}[\|Y_i(\cdot) - Y_i^{(M)}(\cdot)\|_{i,M}^2] + \sum_{j=0}^{N} \mathbb{E}[\|Z_j(\cdot) - Z_j^{(M)}(\cdot)\|_{j,M}^2]\Delta$$

is bounded above by CN^{-1} for an algorithm complexity bounded above by

$$\mathcal{C} = C(N^{3 + \frac{d}{2} + \frac{d}{2(n-1)}} \mathbf{1}_{d \ge 2} + CN^{4 + \frac{d}{2(n-1)}} \mathbf{1}_{d=1}) \log^{d+2}(N).$$

Proof. We use again local polynomials of degree n on hypercubes as the basis functions $p_{0,i}$ and local polynomials of degree $n-1$ on hypercubes as the basis functions $p_{l,i}$. The hypercube lengths are as in Proposition 3.3.10, so that

$$K_{0,i}^{(M)} = CN^{\frac{d}{2}}(N-i)^{-\frac{d(n-1)}{2n}} \log^d(N) \quad \text{and} \quad K_{l,i}^{(M)} = CN^{\frac{dn}{2(n-1)}}(N-i)^{-d/2} \log^d(N).$$

Using the method of Proposition 3.4.21, M_i is set to dominate

$$N^2 \sum_{l=1}^{q} K_{l,i}^{(M)} \log(\varepsilon_Z^{-1}) = CN^{2 + \frac{dn}{2(n-1)}}(N-i)^{-d/2} \log^{1+d}(N)$$

for each $i \in \{0, \ldots, N-1\}$. The highest computational cost of this algorithm again the simulations, which cost a total of $CN \sum_{i=0}^{N-1} M_i$. Therefore, the algorithm complexity is dominated by

$$\begin{aligned}\mathcal{C} &= CN^{3+\frac{dn}{2(n-1)}}(1 + \int_0^{N-1} (N-t)^{-\frac{d}{2}} dt) \log^{1+d}(N) \\ &\leq C(N^{3+\frac{d}{2}+\frac{d}{2(n-1)}} \mathbf{1}_{d\geq 2} + CN^{4+\frac{d}{2(n-1)}} \mathbf{1}_{d=1}) \log^{2+d}(N).\end{aligned}$$

$\square$

Finally, we would like to make a comparison between the adapted LSMDP of this chapter and the LSMDP of Chapter 2. In order to make a fair comparison, we need to make an assumption a further assumption on the bounds of the gradients of the functions $Y_t(\cdot)$ and $Z_t(\cdot)$:

($\mathbf{A}'_{\partial\mathbf{\Phi}}$) The assumption ($\mathbf{A}_{\partial\mathbf{\Phi}}$) holds. The constant C is allowed to have dependencies as in ($\mathbf{A}_{\partial\mathbf{\Phi}}$). Additionally, the functions $Y_t(\cdot)$ are n-times continuously differentiable, and $Z_t(\cdot)$ are $(n-1)$-times continuously differentiable, and there is a constant C such that

$$\left.\begin{array}{rcll} \left|\frac{\partial^r}{\partial x_{i_1}\ldots\partial x_{i_r}} Y_t(x)\right| & \leq & C & \text{for } r = 1, \ldots, n, \\ \\ \left|\frac{\partial^r}{\partial x_{i_1}\ldots\partial x_{i_r}} Z_t(x)\right| & \leq & C & \text{for } r = 1, \ldots, n-1 \end{array}\right\} \quad \text{for all } t \in [0, T) \tag{3.5.4}$$

where $\{i_1, \ldots, i_r\} \in \{1, \ldots, d\}^r$.

The assumption ($\mathbf{A}'_{\partial\mathbf{\Phi}}$) is quite natural. In fact, in the recent work of [CD12, Lemma 3.4], gradient bounds of the form (3.4.2) are obtained under assumptions that the coefficients b and σ of the SDE 3.1.2 are time-homogeneous and the driver f is n-times differentiable; it is also not necessary that σ satisfies ($\mathbf{A}_{\text{u.e.}}$), but a weaker condition. We do not use these results in our analysis, but feel it is useful to mention them to motivate the assumption ($\mathbf{A}'_{\partial\mathbf{\Phi}}$). The following proposition is proved analogously to Proposition 3.5.2 - the gradient bounds (3.5.1) must be updated with (3.5.4), but techniques are otherwise the same - and we do not provide the details. the gradient bounds

Proposition 3.5.3. *Let $n \in \{2, 3, \ldots\}$ and assume ($\mathbf{A}'_{\partial\mathbf{\Phi}}$). There is a choice of basis functions and simulation sizes $(M_0, \ldots, M_N)$ and a constant C (independent of N) with dependencies as in ($\mathbf{A}'_{\partial\mathbf{\Phi}}$) so that, for every N sufficiently large, the error*

$$\max_{0\leq i\leq N-1} \mathbb{E}[\|\bar{y}_i(\cdot) - \bar{y}_i^{(M)}(\cdot)\|_{i,M}^2] + \sum_{j=0}^{N} \mathbb{E}[\|\bar{z}_j(\cdot) - \bar{z}_j^{(M)}(\cdot)\|_{j,M}^2]\Delta$$

is bounded above by CN^{-1} for an algorithm complexity bounded above by

$$\mathcal{C} = CN^{4+\frac{d}{2(n-1)}} \log^{1+d}(N).$$

The complexity of LSMDP computed in Section 2.4.3 is (taking $\theta_{\text{conv}} = 1/2$ and $\kappa + \eta = n - 1$)

$$\mathcal{C}_{LSMDP} = CN^{4+\frac{d}{(n-1)}} \log^{2+2d}(N).$$

Comparing this to the complexity computed in Proposition 3.3.11, one sees that the complexity of

LSMDP is worse in the logarithmic terms and also in the term $N^{d/(n-1)}$.

3.6 Appendix

Lemma 3.6.1 (Conditional expectations)**.** *Suppose that $\mathcal{G}$ and $\mathcal{H}$ are independent sub-σ-algebras of $\mathcal{F}$. Let $F : \Omega\times\mathbb{R}^d \to \mathbb{R}^d$ be bounded and $\mathcal{G}\otimes\mathcal{B}(\mathbb{R}^d)$-measurable, and $U : \Omega \to \mathbb{R}^d$ be $\mathcal{H}$-measurable. Then, $\mathbb{E}[F(U)|\mathcal{H}] = j(U)$ where $j(h) = \mathbb{E}[F(h)]$ for all $h \in \mathbb{R}^d$.*

The proof of Lemma 3.6.1 follows from the Monotone Class Theorem.

Lemma 3.6.2 (Conditional Fubini)**.** *Let $f_s \in \mathcal{H}^2$. Then, for all $t \in [0, T]$, there exists a measurable function*

$$F_t : ([0, T] \times \Omega, \mathcal{B}([0, T]) \otimes \mathcal{F}_t) \to (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$$

belonging to $L_2([0, T] \times \Omega)$ such that $F_t(\cdot, s)$ is a version of $\mathbb{E}_t[f_s]$ for all s and

$$\mathbb{E}_t[\int_0^T f_s ds] = \int_0^T F_t(\cdot, s)ds.$$

Lemma 3.6.2 is proved in Chapter 1. See Lemma 1.8.1.

Lemma 3.6.3 (Conditional Hoeffding inequality)**.** *Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$, and suppose that $X_1, \ldots, X_n$ are $\mathbb{R}$-valued random variables that are "$\mathcal{G}$-conditionally independent" in the sense that, for any subset $\mathcal{Q} \subset \{1, \ldots, n\}$ and bounded measurable functions f_j, $j \in \mathcal{Q}$, the equality*

$$\mathbb{E}[\prod_{j\in\mathcal{Q}} f_j(X_j)|\mathcal{G}] = \prod_{j\in\mathcal{Q}} \mathbb{E}[f_j(x_j)|\mathcal{G}]$$

holds. Additionally, let $X_i \in [a_i, b_i]$ a.s. for some $-\infty < a_i \le b_i < \infty$. Writing $S = X_1+\ldots+X_n$, it follows that, for any $t \ge 0$,

$$\mathbb{P}(|S - \mathbb{E}[S|\mathcal{G}]| \ge t|\mathcal{G}) \le 2\exp\Big(-\frac{2t^2}{\sum_{i=1}^n |b_i - a_i|^2}\Big)$$

The proof of Lemma 3.6.3 is analogous to [GKKW02, Lemma A.3]; one must only replace the independence with conditional independence and the use of the inequality $P(X > \epsilon) \le e^{-t\epsilon}\mathbb{E}[e^{tX}]$ for all $t > 0$ with it's conditional version.

Proposition 3.6.4 (Regression coefficients)**.** *Recall the least-squares regression in Definition 3.3.1. The following properties are satisfied:*

i) Linearity: if $\alpha_1 \in \mathbb{R}^K$ solves $LS\big(X^{1:M}, S_1^{1:M}, p(x)\big)$, $\alpha_2 \in \mathbb{R}^K$ solves $LS\big(X^{1:M}, S_2^{1:M}, p(x)\big)$, and $(\beta_1, \beta_2) \in \mathbb{R}^2$, then $\beta_1\alpha_1 + \beta_2\alpha_2$ solves $LS\big(X^{1:M}, \beta_1 S_1^{1:M} + \beta_2 S_2^{1:M}, p(x)\big)$.

ii) Contraction property: if $\alpha \in \mathbb{R}^K$ solves $LS\big(X^{1:M}, S^{1:M}, p(x)\big)$, then $\frac{1}{M}\sum_{m=1}^M |\alpha \cdot p(X^m)|^2 \le \frac{1}{M}\sum_{m=1}^M |S^m|^2$.

iii) Conditional expectation solution: suppose that $\big(p(X^m)\big)_{m=1,\ldots,M}$ is measurable with respect to the σ-algebra $\mathcal{Q}$ and $\alpha \in \mathbb{R}^K$ solves $LS\big(X^{1:M}, S^{1:M}, p(x)\big)$. Then, denoting by $\mathbb{E}(S^{1:M}|\mathcal{Q})$ the random $\mathbb{R}^M$-vector with m-th entry $(\mathbb{E}(S^m|\mathcal{Q}))$, $\mathbb{E}(\alpha|\mathcal{Q})$ solves $LS\big(X^{1:M}, \mathbb{E}(S^{1:M}|\mathcal{Q}), p(x)\big)$.

The proof of Proposition 3.6.4 is exactly the same as the proof of Lemma 2.4.1 in Chapter 2. In fact, on closer inspection, the two propositions are exactly the same.

4 Approximation of locally Lipschitz BSDEs using Malliavin weights and least-squares regression

4.1 Introduction

Let $T > 0$ be a fixed terminal time and W be a q-dimensional Brownian motion defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$ whose filtration is possibly larger than that generated by the brownian motion. Let $\pi = \{0 =: t_0 < t_1 < \ldots < t_{N-1} < T := t_N\}$ be a given time-grid on $[0, T]$ and $\Delta_i := t_{i+1} - t_i$. Additionally, let $\xi \in L_2(\mathbb{R}; \mathcal{F}_T)$ (i.e. a real-valued and $\mathcal{F}_T$-measurable random variable) and, for $0 \le i < j \le N$, $H^i_j \in \mathbf{L}_2((\mathbb{R}^q)^\top; \mathcal{F}_T)$.

In this chapter, we introduce a numerical algorithm, named **MWLS**, to approximate discrete time stochastic processes (Y, Z) defined by

$$\begin{cases} Y_i & = \mathbb{E}_i[\xi + \sum_{j=i+1}^{N-1} f_j(Y_j, Z_j)\Delta_j], \quad 0 \le i \le N, \\ Z_i & = \mathbb{E}_i[\xi H^i_N + \sum_{j=i+1}^{N-1} f_j(Y_j, Z_j) H^i_j \Delta_j], \quad 0 \le i \le N-1, \end{cases} \tag{4.1.1}$$

where $\mathbb{E}_i[\cdot] = \mathbb{E}[\cdot|\mathcal{F}_{t_i}]$ and $(\omega, y, z) \mapsto f_j(\omega, y, z)$ is $\mathcal{F}_{t_j} \otimes \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}^q)$-measurable.

The main application of (4.1.1) is to approximate continuous time Forward-Backward SDEs (FBSDEs) of the form

$$Y_t = \xi + \int_t^T f(s, X_s, Y_s, Z_s)ds + \int_t^T Z_s dW_s \tag{4.1.2}$$

where $(X_s)_{s\ge 0}$ is a diffusion and ξ is of the form $\Phi(X_T)$ or $\Phi(X_{s_1}, \ldots, X_{s_L})$ $(0 \le s_1 < \ldots < s_L \le T)$, using the discrete time processes. Indeed, according to [MZ02, Theorem 4.2, Theorem 5.1], there is a version of the process $(Z_t)_{0\le t<T}$ given by

$$Z_t = \mathbb{E}_t[\xi H^t_T + \int_t^T f(s, X_s, Y_s, Z_s) H^t_s ds] \tag{4.1.3}$$

where the processes H^s_r are Malliavin weights defined by

$$H^s_r = \frac{1}{r-s}\Big(\int_s^r (\sigma^{-1}(t, X_t) D_s X_t)^\top dW_t\Big)^\top, \quad 0 \le s < r \le T \tag{4.1.4}$$

where $D_s X_t$ is the Malliavin derivative of X_t at s and $\sigma(.)$ is the diffusion coefficient of X. A discretization procedure to approximate the FBSDE (4.1.2) with (4.1.1), including explicit definitions of the random variables H^i_j based on (4.1.4), is given in Chapter 1, where the author also computes the discretization error in terms of N. In honour of the connection between (4.1.1) and (4.1.2), call the random variables H^i_j Malliavin weights, ξ the terminal condition, and $(i, \omega, y, z) \mapsto f_i(y, z)$ the driver. We say that the pair (Y, Z) satisfying (4.1.1) solves a discrete BSDE with data (ξ, f_i).

In this chapter, we are not concerned with the discretization procedure, rather with the analysis of discrete BSDEs (4.1.1) and the algorithm MWLS. Since the system (4.1.1) may be relevant to problems beyond the FBSDE system (4.1.2), we allow the framework and assumptions to accomodate as much generality as possible. However, MWLS is, to the best of our knowledge, the first direct implementation of formula (4.1.3) in a fully implementable numerical scheme.

We revisit the assumptions for the driver and terminal condition of Chapter 2, which are stated in Section 4.2. The driver $f_i(y, z)$ is Lipschitz continuous in (y, z) with a time dependent Lipschitz constant. This allows the algorithm to be used for the numerical approximation of continuous time quadratic BSDEs with bounded, Hölder continuous terminal condition. Furthermore, the driver has a time dependent bound at $(y, z) = (0, 0)$, which allows the use of this method for a certain proxy variance reduction method.

We prove stability results in Proposition 4.3.4. These results are instrumental throughout the chapter. Discrete Gronwall inequalities, outlined in Section 4.3.1, are used to obtain these results. The stability estimates of Z are at the individual time points rather than the time averaged estimates of Proposition 2.3.3 of Chapter 2. This allows for finer and more precise computations. This improvement is coherent with the representation theorem of [MZ02].

We prove almost sure bounds for (Y, Z) depending on the regularity of the terminal condition in Corollary 4.3.5. The almost sure bound for Z_i depends on the remaining time $(T - t_i)$ rather than the time-increment Δ_i, which is similar to the continuous-time case [DG06]. We suitably take advantage of this intrinsic property for the Monte Carlo algorithm MWLS, as we explain in the following paragraphs.

In section 4.3.4, we replace conditional expections of (4.1.1) by $\mathbf{L}_2$-projections, in the style of [BD07] and the MDP scheme (2.1.4) of Chapter 2. The rate of convergence is the same as for both these papers, in that it is the time-averaged projection errors of the discrete BSDE, but the results are in a stronger norm; see Theorem 4.3.6.

Section 4.4 is the core of the chapter and is dedicated to MWLS. In MWLS, the conditional expectations in (4.1.1) are replaced by Monte Carlo least-squares regression. To each point of the time-grid, we associate a cloud of independent sets of independent paths. The algorithm is detailed in Section 4.4.1 and a full error analysis is performed in Section 4.4.2. This error analysis is then used to perform a complexity analysis in Section 4.4.3. We calibrate the numerical parameters to optimize the efficiency of the algorithm.

We assume additional smoothness properties of the solutions of the discrete BSDE to investigate how this improves numerical efficiency. This is relevant in the context of the recent work of [CD12], where the smoothness of the class of continuous time BSDEs that can be approximated by MWLS is investigated.

The complexity analysis is performed with non-uniform time-grids and non-Lipschitz continuous terminal conditions. This is, to the best of our knowledge, novel and relevent to recent works [GM10][GGG12]. We simulate a different number of sample paths at each time point to optimize numerical efficiency: this feature of the scheme is also unique in the literature.

We perform a quantitative comparison to the adapted LSMDP algorithm of Chapter 2 in Section 4.4.3, and find that the MWLS is systematically more efficient, at least in terms of the complexity analysis presented here. An order 1 improvement is observed. This comparison is made in the case of Lipschitz continuous terminal condition and uniform time-grid, which are the assumptions of Chapter 3. In the general setting (non-uniform time-grid $t_i = T - T(1 - i/N)^{1/(\delta\theta_\Phi)}$ and θ_Φ-Hölder continuous terminal condition the complexity of the algorithm is

$$\mathcal{C} = CN^{2+\frac{d}{2n}+1\vee\Psi}\ln^{3+d}(N).$$

for $\Psi := \frac{1}{\theta_\Phi \delta} + d\frac{n+1-\theta_\Phi}{2\theta_\Phi \delta n} - \frac{\theta_c}{\theta_\Phi \delta} \wedge \frac{1}{2\delta}$. Generally, we must choose $\delta \in (0,1)$ [GM10]. We see from this analysis that, not surprisingly, the efficiency of the algorithm deteriorates as the Hölder coefficient deteriorates.

4.2 Assumptions

The following assumptions hold throughout the entirety of the chapter.

$(\mathbf{A}_\xi)$ ξ is a bounded $\mathcal{F}_T$-measurable random variable, with

$$C_\xi := \mathbb{P} - \operatorname{ess\,sup}_\omega |\xi(\omega)| < +\infty.$$

$(\mathbf{A_F})$ i) $(\omega, y, z) \mapsto f_i(y,z)$ is $\mathcal{F}_{t_i} \otimes \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}^q)$-measurable for every $i < N$, and there exist deterministic parameters $\theta \in (0,1]$ and $L_f \in [0,+\infty)$ such that

$$|f_i(y,z) - f_i(y',z')| \leq \frac{L_f}{(T-t_i)^{(1-\theta)/2}}(|y-y'| + |z-z'|),$$

for any $(y, y', z, z') \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^q \times \mathbb{R}^q$.

ii) There exist deterministic parameters $\theta_c \in (0,1]$ and $C_f \in [0,+\infty)$ such that

$$|f_i(0,0)| \leq \frac{C_f}{(T-t_i)^{1-\theta_c}}, \quad \forall 0 \leq i < N.$$

iii) There exists $R_\pi > 0$ such that the time-grids $\pi := \{0 = t_0 < \ldots < t_N = T\}$ satisfy

$$\limsup_{N\to\infty} \max_{0\leq i \leq N-2} \frac{\Delta_{i+1}}{\Delta_i} \leq R_\pi$$

$(\mathbf{A_H})$ For all $0 \leq i < j \leq N$, the Malliavin weights satisfy

$$\mathbb{E}[H^i_j | \mathcal{F}_{t_i}] = 0, \quad \left[\mathbb{E}[|H^i_j|^2 | \mathcal{F}_{t_i}]\right]^{1/2} \leq \frac{C_M}{(t_j - t_i)^{1/2}}.$$

We remark that assumptions $(\mathbf{A_F})$and $(\mathbf{A}_\xi)$ are the almost same as their equivalents in Section 2.2 of Chapter 2, but that $(\mathbf{A_F}$-$\mathbf{iii})$ has been relaxed: it clearly admits the time-grids of [Ric11], and may be valuable for future work on time-grid optimization.

The following assumptions will be necessary for Section 4.4.

$(\mathbf{A_X})$ X is a Markov chain in $\mathbb{R}^d$ $(1 \leq d < +\infty)$ adapted to $(\mathcal{F}_{t_i})_i$. For every $i < N$ and $j > i$, there exist $\mathcal{G}_i \otimes \mathcal{B}(\mathbb{R}^d)$-measurable functions $V^i_j : \Omega \times \mathbb{R}^d \to \mathbb{R}^d$ where $\mathcal{G}_i$ is independent of $\sigma(X_i)$, such that $X_j = V^i_j(X_i)$.

$(\mathbf{A}'_\xi)$ ξ is of form $\xi = \Phi(X_N)$ for a measurable function Φ.

$(\mathbf{A'_F})$ For every $i < N$, the driver is of the form $f_i(\omega, y, z) = f_i(X_i(\omega), y, z)$, and $(x, y, z) \mapsto f_i(x, y, z)$ is $\mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}^q)$-measurable and $(\mathbf{A_F})$ is satisfied.

($\mathbf{A'_H}$) For every $i < N$ and $j > i$, there is a function $h^i_j : \Omega \times \mathbb{R}^d \to \mathbb{R}^q$ that is $\mathcal{G}_i \otimes \mathcal{B}(\mathbb{R}^d)$-measurable, where $\mathcal{G}_i$ is independent of $\sigma(X_i)$, such that $H^i_j = h^i_j(X_i)$ almost surely.

These give us a Markov representation for solutions of the discrete BSDEs:

Lemma 4.2.1. *Assume* ($\mathbf{A_X}$), ($\mathbf{A'_\xi}$), ($\mathbf{A'_F}$)*and* ($\mathbf{A'_H}$)*. For all $i < N$, there exist measurable, deterministic functions $y_i : \mathbb{R}^d \to \mathbb{R}$ and $z_i : \mathbb{R}^d \to \mathbb{R}^q$ such that $Y_i = y_i(X_i)$ and $Z_i = z_i(X_i)$ holds almost surely. Moreover,*

$$\left.\begin{array}{ll} y_i(x) & = \int (\Phi(x_N) + \sum_{j=i+1}^{N-1} f_j(x_j, y_j(x_j), z_j(x_j))\Delta_j) d\mu^x_i \\ z_i(x) & = \int (\Phi(x_N)h_N + \sum_{j=i+1}^{N-1} f_j(x_j, y_j(x_j), z_j(x_j))h_j\Delta_j) d\lambda^x_i \end{array}\right\} \tag{4.2.1}$$

*where μ^x_i is the joint law of $(X_{i+1}, \dots, X_N)$ given that $X_i = x$, and λ^x_i is the joint law of $(X_{i+1}, \dots, X_N,$
$H^i_{i+1}, \dots, H^i_N)$ conditional on $X_i = x$.*

We apply the following lemma to prove Lemma 4.2.1:

Lemma 4.2.2. *Let $F : \Omega \times \mathbb{R}^d \to \mathbb{R}^d$ be bounded and $\mathcal{G} \otimes \mathcal{B}(\mathbb{R}^d)$-measurable, and $U : \Omega \to \mathbb{R}^d$ be $\mathcal{H}$-measurable. Then, $\mathbb{E}[F(U)|\mathcal{H}] = j(U)$ where $j(h) = \mathbb{E}[F(h)]$ for all $h \in \mathbb{R}^d$.*

Lemma 4.2.2 follows from the Monotone Class Theorem.

Proof. ***(of Lemma 4.2.1)*** Due to ($\mathbf{A_X}$), ($\mathbf{A'_F}$) and ($\mathbf{A'_H}$), there exist functions $V^i_j : \Omega \times \mathbb{R}^d \to \mathbb{R}^d$ and $h^i_j : \Omega \times \mathbb{R}^d \to \mathbb{R}^q$ that are $\mathcal{G}_i \otimes \mathcal{B}(\mathbb{R}^d)$-measurable, where $\mathcal{G}_i$ is independent of $\sigma(X_i)$, such that $X_j = V^i_j(X_i)$ and $H^i_j = h^i_j(X_i)$ almost surely. The Markovian form and (4.2.1) then follow by using the linearity of conditional expectations and Lemma 4.2.2 in (4.1.1). □

4.2.1 Applications

We outline two canonical examples to motive the assumptions in ($\mathbf{A_F}$). Take $\xi = \Phi(X_T)$ and $f(t, \omega, y, z) = f(t, X_t(\omega), y, z)$ where X is a Brownian diffusion. For simplicity, assume $q = d$ and that the coefficients of the diffusion X are smooth and bounded and that its diffusion coefficient $\sigma(t, x)$ satisfies a uniform ellipticity condition.

Quadratic BSDEs. We denote by $\mathcal{L}$ the infinitesimal generator of X. Consider a quadratic growth driver satisfying

$$|f(t, x, y, z)| \leq c\ (1 + |y| + |z|^2),$$
$$|f(t, x, y, z) - f(t, x, y', z')| \leq c\ (1 + |z| + |z'|)(|x - x'| + |y - y'| + |z - z'|)$$

for any $(t, x, x', y, y', z, z') \in [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d$ and for a given constant $c \geq 0$. Assume additionally that the terminal function Φ is Hölder continuous and bounded. Then [DG06, Theorem 2.1] yields that the continuous-time BSDE is given by $Y_t = u(t, X_t)$ and $Z_t = \nabla u(t, X_t)\sigma(t, X_t)$ where u solves the semi-linear PDE $\partial_t u(t, x) + \mathcal{L}u(t, x) + f(t, x, u(t, x), \nabla u(t, x)\sigma(t, x)) = 0$ with $u(T, x) = \Phi(x)$. Moreover, there exist constants $\theta \in (0, 1]$ and $C_u \in \mathbb{R}^+$ such that

$$(T - t)^{(1-\theta)/2} |\nabla u(t, x)\sigma(t, x)| \leq C_u, \quad \forall (t, x) \in [0, T) \times \mathbb{R}^d.$$

Now, set $\varphi_t : \zeta \in \mathbb{R} \mapsto \varphi_t(\zeta) = \text{sign}(\zeta) \min\big(|\zeta|, \frac{C_u}{(T-t)^{(1-\theta)/2}}\big)$ and define the new driver $\bar{f}(t,x,y,z) := f\big(t,x,y,\varphi_t(z_1)\ \ldots, \varphi_t(z_d)\big)$. Observe that $\bar{f}(t,X_t,Y_t,Z_t) = f(t,X_t,Y_t,Z_t)$, thus it is equivalent to solve the BSDE with driver f or $\bar{f}$. Notice also that $\varphi_t(\cdot)$ is 1-Lipschitz continuous and bounded by $\frac{C_u}{(T-t)^{(1-\theta)/2}}$, hence $f_i(y,z) := \bar{f}(t_i, X_{t_i}, y, z)$ satisfies ($\mathbf{A_F}$**-i-ii**)with $C_f = c$, $\theta_c = 1$, $L_f = c(T^{(1-\theta)/2} + 2\sqrt{d}C_u)$, $\theta_L = \theta$. In Chapter 1, it was proved that the exponent θ is equal to the Hölder exponent of Φ, and it is possible to obtain an explicit estimate of the constant C_u.

What is particularly interesting is that the computational efficiency computed in Section 4.4.3 is of the same order whether we take local or global Lipschitz continuity of the driver. The local continuity corresponds to the quadratic BSDE problem. In this sense, we obtain the optimal rates of convergence for the quadratic BSDE. The truncation method of [IDR10, Section 6], for example, requires the use of a truncation of the driver by a smooth projection of the z component onto the open ball or radius $\bar{R}$. This will reduce the quadratic driver to a Lipschitz continuous driver whose Lipschitz coefficent depends on $\bar{R}$, and they require that $\bar{R}$ be large so that their approximation error is low [IDR10, Theorem 6.2]. As we will see in Theorems 2.4.4 and 2.4.5, the efficiency of the LSMDP algorithm depends on the absolute bounds of y_i and z_i, which depend exponentially on the Lipschitz constant, see Proposition 2.3.8. This presents a substantial reduction of the numerical efficiency of the numerical scheme. This phenomenon has already been observed in the introduction of [Ric11], where a nice heuristic argument is provided. Another alternative approach is to apply he Cole-Hopf transform [IDRZ10], which transforms a quadratic BSDE into a Lipschitz BSDE. Numerical resolution of the transformed problem leads to good convergence rates. However, the Cole-Hopf transform can only be applied to certain quadratic drivers. Our method accommodates general quadratic drivers. The method of [Ric11] the use of an irregular time-grid to attain an optimal convergence rate. Although our algorithm accommodates this special time grid, it is also able to accommodate the time grids of Chapter 1, where the time points take the form $t_i = T - T(1 - i/N)^{1/\theta_\pi}$. It was shown in Chapter 1 that these time grids improve the rate of convergence of the time-discrete scheme to the continuous BSDE.

Using proxys for numerical stability. Consider the case $\theta_L = 1$, and assume we explicitly know $(v(t,x), \nabla v(t,x)\sigma(t,x))$, the solution to the linear parabolic equation $\partial_t v(t,x) + \tilde{\mathcal{L}}v(t,x) + \tilde{f}(t,x) = 0$; the diffusion process associated to $\tilde{\mathcal{L}}$, the terminal condition and the driver may have changed to produce an analytical solution. v is called *proxy* in [BGM09]. It is then natural to numerically compute the residual $(Y^0_t, Z^0_t) := (Y_t - v(t,X_t), Z_t - \nabla v(t,X_t)\sigma(t,X_t))$. It solves a BSDE with terminal function $\Phi(.) - v(T,.)$ and driver

$$f^0(t,x,y,z) := f(t,x,y + v(t,x), z + \nabla v(t,x)\sigma(t,x)) - \tilde{f}(t,x) + (\mathcal{L} - \tilde{\mathcal{L}})v(t,x).$$

The new driver f^0 is uniformly Lipschitz w.r.t. y and z. If $v(T,.)$ is θ_Φ-Hölder continuous ($\theta_\Phi \in (0,1]$), then usual PDE estimates on the parabolic operator $\tilde{\mathcal{L}}$ give $\sup_{t<T}(T-t)^{(\frac{k-\theta_\Phi}{2})_+}|\nabla_x^{(k)} v(t,x)| \leq C_v$ $(k = 0,1,2)$, from which (1.1.3) is satisfied for f^0 with $\theta_c = \frac{1+\theta_\Phi}{2}$.

To complete this example, we mention that in the case $\tilde{\mathcal{L}} = \mathcal{L}$, $v(T,.) = \Phi(.)$ and $\tilde{f} = 0$. This has been investigated numerically in, for example, [BS12], where they include the solution $v(t,\cdot)$ as part of the numerical approximation of (Y^0, Z^0) and demonstrate that this may be more efficient than using a naïve method.

In general, even when no proxy is available, it is still useful to use the decomposition $v(T,\cdot) =$

$\Phi(\cdot)$, $\tilde{\mathcal{L}} = \mathcal{L}$ of the BSDE and to approximate $v(t,\cdot)$ and $\nabla v(t,\cdot)$ at $t < T$. The numerical schemes for (Y^0, Z^0) are much better behaved due to zero terminal condition; this phenomenon is investigated in Chapter 3. Moreover, approximation of $(v(\cdot), \nabla v(\cdot))$ can be performed using a multilevel technique; see Chapter 3. This technique incorporates variance reduction and can be performed using parallel computing, leading to high efficiency of the overall method.

4.3 Stability

Suppose that (Y_1, Z_1) (resp. (Y_2, Z_2)) solves a discrete BSDE with data $(\xi_1, f_{1,i})$ (resp. $(\xi_2, f_{2,i})$). We are mainly interested in studying the differences $(Y_1 - Y_2, Z_1 - Z_2)$.

4.3.1 Gronwall type inequalities

Here we gather deterministic inequalities frequently used throughout the chapter. They show how linear inequalities with singular coefficients propagate. They take the form of unusual Gronwall type inequalities. Their proofs are postponed to Appendix 4.5.

Lemma 4.3.1. *Let $R_\pi \geq 1$ and $\alpha, \beta > 0$. For π in the class of time-grids satisfying $(\mathbf{A_F})$(iii) there exists a constant $B_{\alpha,\beta}$ depending only on R_π, α and β (but not on the time-grid) such that, for any $0 \leq i < k \leq N$,*

$$\sum_{j=i+1}^{k-1} (t_k - t_j)^{\alpha-1}(t_j - t_i)^{\beta-1}\Delta_j \leq B_{\alpha,\beta}(t_k - t_i)^{\alpha+\beta-1}.$$

Lemma 4.3.2 (exponent improvment in recursive equations)**.** *Let $\alpha \geq 0, \beta > 0$ and $k \in \{0, \dots, N-1\}$. Suppose that, for a positive constant C_u, the finite positive real numbers $\{u_l\}_{l\geq k}$ and $\{w_l\}_{l\geq k}$ satisfy*

$$u_j \leq w_j + C_u \sum_{l=j+1}^{N-1} \frac{u_l \Delta_l}{(T - t_l)^{\frac{1}{2}-\beta}(t_l - t_j)^{\frac{1}{2}-\alpha}}. \tag{4.3.1}$$

Then, for two constants $\mathcal{C}_{(4.3.2a)}$ and $\mathcal{C}_{(4.3.2b)}$ that depend only on C_u, T, α, β and R_π,

$$u_j \leq \mathcal{C}_{(4.3.2a)} w_j + \mathcal{C}_{(4.3.2a)} \sum_{l=j+1}^{N-1} \frac{w_l \Delta_l}{(T - t_l)^{\frac{1}{2}-\beta}(t_l - t_j)^{\frac{1}{2}-\alpha}} + \mathcal{C}_{(4.3.2b)} \sum_{l=j+1}^{N-1} \frac{u_l \Delta_l}{(T - t_l)^{\frac{1}{2}-\beta}}. \tag{4.3.2}$$

Lemma 4.3.3 (a priori estimates)**.** *Let $\alpha \geq 0, \beta > 0$ and $k \in \{0, \dots, N-1\}$. Assume that the finite positive real numbers $\{u_l\}_{l\geq k}$ and $\{w_l\}_{l\geq k}$ satisfy* (4.3.2) *for some positive constants $\mathcal{C}_{(4.3.2a)}$ and $\mathcal{C}_{(4.3.2b)}$. Then, for $\gamma > 0$, there is a positive constant $\mathcal{C}^{(\gamma)}_{(4.3.3)}$ (depending only on $\mathcal{C}_{(4.3.2a)}$, $\mathcal{C}_{(4.3.2b)}$, T, α, β,and γ) such that*

$$\sum_{l=j+1}^{N-1} \frac{u_l \Delta_l}{(T - t_l)^{\frac{1}{2}-\beta}(t_l - t_j)^{1-\gamma}} \leq \mathcal{C}^{(\gamma)}_{(4.3.3)} \sum_{l=j+1}^{N-1} \frac{w_l \Delta_l}{(T - t_l)^{\frac{1}{2}-\beta}(t_l - t_j)^{1-\gamma}}. \tag{4.3.3}$$

4.3.2 Stability of discrete BSDEs

Define:

$$\Delta Y = Y_1 - Y_2, \quad \Delta Z = Z_1 - Z_2, \quad \Delta\xi = \xi_1 - \xi_2,$$
$$\Delta f_i = f_{1,i}(Y_{1,i}, Z_{1,i}) - f_{2,i}(Y_{1,i}, Z_{1,i}).$$

To simplify notation, we write $|X|_{2,k} = \sqrt{\mathbb{E}_k[X^2]}$ for any square integrable random variable X, and, for $j \geq i$, define

$$|\Theta_j|_{2,k} = |\Delta Y_j|_{2,k} + |\Delta Z_j|_{2,k}.$$

Using $(\mathbf{A_H})$, we obtain $\mathbb{E}_i[\Delta\xi H_N^i] = \mathbb{E}_i[(\Delta\xi - \mathbb{E}_i\Delta\xi)H_N^i]$. Combining this with $(\mathbf{A_F}\text{-}\mathbf{i})$, our stability equations (for $k \leq i$) are

$$|\Delta Y_i|_{2,k} \leq |\Delta\xi|_{2,k} + \sum_{j=i+1}^{N-1} |\Delta f_j|_{2,k}\Delta_j + \sum_{j=i+1}^{N-1} \frac{L_{f_2}|\Theta_j|_{2,k}}{(T-t_j)^{(1-\theta)/2}}\Delta_j, \tag{4.3.4}$$

$$\begin{aligned} |\Delta Z_i|_{2,k} \leq{} & \frac{C_M|\Delta\xi - \mathbb{E}_i\Delta\xi|_{2,k}}{\sqrt{T-t_i}} + \sum_{j=i+1}^{N-1} \frac{C_M|\Delta f_j|_{2,k}}{\sqrt{t_j - t_i}}\Delta_j \\ & + \sum_{j=i+1}^{N-1} \frac{L_{f_2}C_M|\Theta_j|_{2,k}}{(T-t_j)^{(1-\theta)/2}\sqrt{t_j-t_i}}\Delta_j, \end{aligned} \tag{4.3.5}$$

Proposition 4.3.4. *Under* $(\mathbf{A_F})$*and* $(\mathbf{A_H})$*, there exist constants* $C_y^{(1)}, C_y^{(2)}, C_z^{(1)}, C_z^{(2)}, C_z^{(3)} \geq 0$ *such that*

$$|\Delta Y_i|_{2,k} \leq C_y^{(1)}|\Delta\xi|_{2,k} + C_y^{(2)} \sum_{j=i+1}^{N-1} |\Delta f_j|_{2,k}\Delta_j,$$

$$|\Delta Z_i|_{2,k} \leq C_z^{(1)}\frac{|\Delta\xi - \mathbb{E}_i\Delta\xi|_{2,k}}{\sqrt{T-t_i}} + C_z^{(2)} \sum_{j=i+1}^{N-1} \frac{|\Delta f_j|_{2,k}}{\sqrt{t_j-t_i}}\Delta_j + C_z^{(3)}|\Delta\xi|_{2,k}(T-t_i)^{\frac{\theta}{2}}.$$

The constants depend only on C_M, L_{f_2}, θ and T.

Proof. Using (4.3.4) and (4.3.5), we obtain

$$\begin{aligned} |\Theta_j|_{2,k} \leq{} & C_M\frac{|\Delta\xi - \mathbb{E}_j\Delta\xi|_{2,k}}{\sqrt{T-t_j}} + |\Delta\xi|_{2,k} + (C_M + \sqrt{T}) \sum_{l=j+1}^{N-1} \frac{|\Delta f_l|_{2,k}\Delta_l}{\sqrt{t_l - t_j}} \\ & + (C_M + \sqrt{T}) \sum_{l=j+1}^{N-1} \frac{L_{f_2}|\Theta_l|_{2,k}\Delta_l}{(T-t_l)^{(1-\theta)/2}\sqrt{t_l-t_j}}. \end{aligned} \tag{4.3.6}$$

In the following, we write $U \leq_c V$ to mean $U \leq C(C_M, L_{f_2}, \theta, T)V$, and $U =_c V$ means $U = C(C_M, L_{f_2}, \theta, T)V$.

Upper bound for (4.3.6). We apply Lemmas 4.3.2 and 4.3.3 under the setting $u_j = |\Theta_j|_{2,k}$, $w_j = C_M\frac{|\Delta\xi - \mathbb{E}_j\Delta\xi|_{2,k}}{\sqrt{T-t_j}} + |\Delta\xi|_{2,k} + (C_M + \sqrt{T})\sum_{l=j+1}^{N-1}\frac{|\Delta f_l|_{2,k}\Delta_l}{\sqrt{t_l - t_j}}$, $\alpha = 0$, $\beta = \frac{\theta}{2}$, $C_u = L_{f_2}(C_M + \sqrt{T})$.

To make results fully explicit, we first need to upper bound quantities of the form

$$\mathcal{I}_{j+1}^{(\gamma)} := \sum_{l=j+1}^{N-1} \frac{w_l \Delta_l}{(T-t_l)^{\frac{1}{2}-\frac{\theta}{2}}(t_l-t_j)^{1-\gamma}}.$$

Using that $|\Delta\xi - \mathbb{E}_l\Delta\xi|_{2,k}$ is non-increasing in l and Lemma 4.3.1, we obtain

$$\begin{aligned}
\mathcal{I}_{j+1}^{(\gamma)} &= \sum_{l=j+1}^{N-1} \frac{C_M \frac{|\Delta\xi-\mathbb{E}_l\Delta\xi|_{2,k}}{\sqrt{T-t_l}} + |\Delta\xi|_{2,k} + (C_M+\sqrt{T})\sum_{r=l+1}^{N-1}\frac{|\Delta f_r|_{2,k}\Delta_r}{\sqrt{t_r-t_l}}}{(T-t_l)^{\frac{1}{2}-\frac{\theta}{2}}(t_l-t_j)^{1-\gamma}}\Delta_l \\
&\leq C_M B_{\frac{\theta}{2},\gamma}\frac{|\Delta\xi-\mathbb{E}_{j+1}\Delta\xi|_{2,k}}{(T-t_j)^{1-\frac{\theta}{2}-\gamma}} + B_{\frac{1}{2}+\frac{\theta}{2},\gamma}\frac{|\Delta\xi|_{2,k}}{(T-t_j)^{\frac{1}{2}-\frac{\theta}{2}-\gamma}} \\
&\quad + (C_M+\sqrt{T})B_{\frac{\theta}{2},\gamma}\sum_{l=j+2}^{N-1}\frac{|\Delta f_l|_{2,k}\Delta_l}{(t_l-t_j)^{1-\frac{\theta}{2}-\gamma}}\Big(=_c w_{j+1}\Big). \qquad (4.3.7)
\end{aligned}$$

Then, applying Lemmas 4.3.2 and 4.3.3, we obtain

$$\begin{aligned}
|\Theta_j|_{2,k} &\leq \mathcal{C}_{(4.3.2a)}\left[C_M\frac{|\Delta\xi-\mathbb{E}_j\Delta\xi|_{2,k}}{\sqrt{T-t_j}} + |\Delta\xi|_{2,k} + (C_M+\sqrt{T})\sum_{l=j+1}^{N-1}\frac{|\Delta f_l|_{2,k}\Delta_l}{\sqrt{t_l-t_j}}\right] \\
&\quad + \mathcal{C}_{(4.3.2a)}\mathcal{I}_{j+1}^{(\frac{1}{2}+\alpha)} + \mathcal{C}_{(4.3.2b)}\mathcal{C}_{(4.3.3)}^{(1)}\mathcal{I}_{j+1}^{(1)} \\
&\leq_c \frac{|\Delta\xi-\mathbb{E}_j\Delta\xi|_{2,k}}{\sqrt{T-t_j}} + |\Delta\xi|_{2,k} + \sum_{l=j+1}^{N-1}\frac{|\Delta f_l|_{2,k}\Delta_l}{\sqrt{t_l-t_j}}\Big(=_c w_j\Big).
\end{aligned}$$

Upper bound for (4.3.4). Starting from (4.3.4), applying Lemma 4.3.3 and using $|\Delta\xi - \mathbb{E}_i\Delta\xi|_{2,k} \leq 2|\Delta\xi|_{2,k}$ and the estimate (4.3.7), we get

$$\begin{aligned}
|\Delta Y_i|_{2,k} &\leq |\Delta\xi|_{2,k} + \sum_{j=i+1}^{N-1}|\Delta f_j|_{2,k}\Delta_j + L_{f_2}\mathcal{C}_{(4.3.3)}^{(1)}\mathcal{I}_{i+1}^{(1)} \\
&\leq_c |\Delta\xi|_{2,k} + \sum_{j=i+1}^{N-1}|\Delta f_j|_{2,k}\Delta_j.
\end{aligned}$$

Upper bound of (4.3.5). Starting from (4.3.5), applying Lemma 4.3.3 and using the estimate (4.3.7), we obtain

$$\begin{aligned}
|\Delta Z_i|_{2,k} &\leq \frac{C_M|\Delta\xi-\mathbb{E}_i\Delta\xi|_{2,k}}{\sqrt{T-t_i}} + \sum_{j=i+1}^{N-1}\frac{C_M|\Delta f_j|_{2,k}}{\sqrt{t_j-t_i}}\Delta_j + L_{f_2}C_M\mathcal{C}_{(4.3.3)}^{(\frac{1}{2})}\mathcal{I}_{i+1}^{(\frac{1}{2})} \\
&\leq_c \frac{|\Delta\xi-\mathbb{E}_i\Delta\xi|_{2,k}}{\sqrt{T-t_i}} + \sum_{j=i+1}^{N-1}\frac{|\Delta f_j|_{2,k}}{\sqrt{t_j-t_i}}\Delta_j + (T-t_i)^{\frac{\theta}{2}}|\Delta\xi|_{2,k}.
\end{aligned}$$

$\square$

4.3.3 Almost sure bounds

Corollary 4.3.5. *Under* $(\mathbf{A}_\xi)$, $(\mathbf{A_F})$*and* $(\mathbf{A_H})$

$$|Y_i| \le C_{y,i} := C_y^{(1)} C_\xi + C_y^{(2)} C_f B_{\theta_c,1} (T-t_i)^{\theta_c}, \tag{4.3.8}$$

$$|Z_i| \le C_{z,i} := C_z^{(1)} \operatorname*{ess\,sup}_\omega \frac{|\xi - \mathbb{E}_i \xi|_{2,i}}{\sqrt{T-t_i}} + \frac{C_z^{(2)} C_f B_{\theta_c,\frac{1}{2}}}{(T-t_i)^{\frac{1}{2}-\theta_c}} + C_z^{(3)} C_\xi (T-t_i)^{\frac{\theta}{2}}. \tag{4.3.9}$$

The above upper bounds are able to handle terminal values ξ under the full generality admitted by $(\mathbf{A}_\xi)$. Without any further information on ξ, we can derive the simple bounds

$$|Y_i| + \sqrt{T-t_i}|Z_i| \le C_{y,z} \tag{4.3.10}$$

for an explicit, time uniform constant $C_{y,z}$. it may, however, be useful to take advantage of additional information on ξ. In Section 4.4.3, we tune the parameters of the MWLS method, and here finer estimates on $C_{y,i}$ and $C_{z,i}$ are useful. Two situations are important.

- For zero terminal condition, Y and Z get smaller and smaller as t_i goes to T as expected: $|Y_i| + \sqrt{T-t_i}|Z_i| \le C(T-t_i)^{\theta_c}$ for a constant C depending only on $C_z^{(2)}$, C_f and θ_c. This result is useful for variance reduction methods like the proxy method of Section 1.3.2, the method of Martingale basis [BS12], and the multilevel method of Chapter 3.
- Under $(\mathbf{A}'_\xi)$, i.e. $\xi = \Phi(X_N)$, and assuming a θ_Φ-Hölder terminal function Φ and $\mathbb{E}_i[|X_N - X_i|^2] \le C_X(T-t_i)$ (satisfied for a diffusion process with bounded coefficients or for its Euler scheme), we obtain an improved estimate for Z: $|Z_i| \le C(T-t_i)^{-\frac{1}{2}+\theta_c \wedge \frac{\theta_\Phi}{2}}$ for some constant C depending only on $C_z^{(1)}$, $C_z^{(2)}$, $C_z^{(3)}$, C_f, θ_c, C_X and the Hölder constant of Φ.

This is why in the subsequent analysis, we keep track on the general dependence on i of the constants $C_{y,i}$ and $C_{z,i}$.

Proof of Corollary 4.3.5. $(0,0)$ is the solution of the discrete BSDE with data $(\xi_1 \equiv 0, f_{1,i} \equiv 0)$. Applying Proposition 4.3.4 with $(Y_1, Z_1) = (0,0)$ and $(Y_2, Z_2) = (Y, Z)$ yields

$$|Y_i|_{2,k} \le C_y^{(1)} |\xi|_{2,k} + C_y^{(2)} \sum_{j=i+1}^{N-1} |f_j(0,0)|_{2,k} \Delta_j,$$

$$|Z_i|_{2,k} \le \frac{C_z^{(1)} |\xi - \mathbb{E}_i \xi|_{2,k}}{\sqrt{T-t_i}} + C_z^{(2)} \sum_{j=i+1}^{N-1} \frac{|f_j(0,0)|_{2,k} \Delta_j}{\sqrt{t_j - t_i}} + C_z^{(3)} |\Delta \xi|_{2,k} (T-t_i)^{\frac{\theta}{2}}.$$

for $i = 0, \dots, N-1$. Taking $k = i$ and plugging in the almost sure bounds on $|\xi|$ $(\mathbf{A}_\xi)$, $|f_j(0,0)|$ $(\mathbf{A_F}$-**ii**) and using Lemma 4.3.1 then yields the result. ☐

4.3.4 Projection errors

For $i = 0, \dots, N-1$, let $\mathcal{S}_i^Y, \mathcal{S}_i^{Z,1}, \dots, \mathcal{S}_i^{Z,q}$ be convex subsets of $\mathbf{L}_2(\mathcal{F}_{t_i}; \mathbb{R})$, and $\mathcal{P}_i^Y, \mathcal{P}_i^{Z,1}, \dots, \mathcal{P}_i^{Z,q}$ be their respective projection operators. In what follows, we write

$$z \in (\mathbb{R}^q)^\top \mapsto \mathcal{P}_i^Z z = (\mathcal{P}_i^{Z,1} z_1, \dots, \mathcal{P}_i^{Z,q} z_q).$$

Using the projection operators in place of the conditional expectations, we define the following approximation $(\hat{Y}, \hat{Z})$ of the discrete BSDE (Y, Z) with data $(\xi, f_j(y, z))$:

$$\hat{Y}_i := \mathcal{P}_i^Y(\xi + \sum_{j=i+1}^{N-1} f_j(\hat{Y}_j, \hat{Z}_j)\Delta_j), \tag{4.3.11}$$

$$\hat{Z}_{l,i} := \mathcal{P}_i^{Z,l}(\xi H_{l,N}^i + \sum_{j=i+1}^{N-1} f_j(\hat{Y}_j, \hat{Z}_j) H_{l,j}^i \Delta_j) \tag{4.3.12}$$

where $H_{l,j}^i$ is the l-th component of the vector H_j^i and $\hat{Z}_{l,i}$ is the l-th component of $\hat{Z}_i$.

Theorem 4.3.6. *There exists a constant $\hat{C}$ independent of the time-grid such that*

$$|Y_i - \hat{Y}_i|_2 \leq |Y_i - \mathcal{P}_i^Y Y_i|_2 + \hat{C} \sum_{j=i+1}^{N-1} \frac{\{|Y_j - \mathcal{P}_j^Y Y_j|_2 + |Z_j - \mathcal{P}_j^Z Z_j|_2\}\Delta_j}{(T - t_j)^{(1-\theta)/2}}, \tag{4.3.13}$$

$$|Z_i - \hat{Z}_i|_2 \leq |Z_i - \mathcal{P}_i^Z Z_i|_2 + \hat{C} \sum_{j=i+1}^{N-1} \frac{\{|Y_j - \mathcal{P}_j^Y Y_j|_2 + |Z_j - \mathcal{P}_j^Z Z_j|_2\}\Delta_j}{(T - t_j)^{(1-\theta)/2}\sqrt{t_j - t_i}} \tag{4.3.14}$$

for all i.

Proof. Define the intermediate discrete BSDE $(\bar{Y}, \bar{Z})$ with data $(\bar{\xi} \equiv \xi, \bar{f}_j(y, z) \equiv f_j(\hat{Y}_j, \hat{Z}_j))$. We observe, from Lemma 2.3.5(a) in Chapter 2, that $\hat{Y}_i = \mathcal{P}_i^Y(\bar{Y}_i)$ and $\hat{Z}_i = \mathcal{P}_i^Z(\bar{Z}_i)$. Using this result, Lemma 2.3.5(b) , Proposition 4.3.4 and the Lipschitz continuity of f_j, we obtain

$$\begin{aligned} |Y_i - \hat{Y}_i|_2 &\leq |Y_i - \mathcal{P}_i^Y Y_i|_2 + |\mathcal{P}_i^Y Y_i - \hat{Y}_i|_2 \leq |Y_i - \mathcal{P}_i^Y Y_i|_2 + |Y_i - \bar{Y}_i|_2 \\ &\leq |Y_i - \mathcal{P}_i^Y Y_i|_2 + C_y^{(2)} \sum_{j=i+1}^{N-1} \frac{L_f\{|Y_j - \hat{Y}_j|_2 + |Z_j - \hat{Z}_j|_2\}\Delta_j}{(T - t_j)^{(1-\theta)/2}} \end{aligned} \tag{4.3.15}$$

Similarly, and introducing the notation $|\Theta_j|_2 := |Y_j - \hat{Y}_j|_2 + |Z_j - \hat{Z}_j|_2$, we obtain

$$|Z_i - \hat{Z}_i|_2 \leq |Z_i - \mathcal{P}_i^Z Z_i|_2 + C_z^{(2)} \sum_{j=i+1}^{N-1} \frac{L_f |\Theta_j|_2 \Delta_j}{(T - t_j)^{(1-\theta)/2}\sqrt{t_j - t_i}} \tag{4.3.16}$$

From now on, $c > 0$ is a constant that may change from line to line, but is always independent of the time-grid. Putting (4.3.15) and (4.3.16) together, we obtain

$$|\Theta_i|_2 \leq c|Y_i - \mathcal{P}_i^Y Y_i|_2 + c|Z_i - \mathcal{P}_i^Z Z_i|_2 + c \sum_{j=i+1}^{N-1} \frac{|\Theta_j|_2 \Delta_j}{(T - t_j)^{(1-\theta)/2}\sqrt{t_j - t_i}} \tag{4.3.17}$$

We now apply Lemmas 4.3.2 in the setting $u_j = \Theta_j$, $w_j = c|Y_j - \mathcal{P}_j^Y Y_j|_2 + c|Z_j + \mathcal{P}_j^Z Z_j|_2$, $C_u = c$, $\beta = \frac{\theta}{2}$, and $\alpha = 0$ to obtain

$$\begin{aligned}|\Theta_i|_2 \leq c|Y_i - \mathcal{P}_i^Y Y_i|_2 + c|Z_i - \mathcal{P}_i^Z Z_i|_2 + c \sum_{j=i+1}^{N-1} \frac{|\Theta_j|_2 \Delta_j}{(T-t_j)^{(1-\theta)/2}} \\ + c \sum_{j=i+1}^{N-1} \frac{\{|Y_j - \mathcal{P}_j^Y Y_j|_2 + |Z_j - \mathcal{P}_j^Z Z_j|_2\}\Delta_j}{(T-t_j)^{(1-\theta)/2}\sqrt{t_j - t_i}}\end{aligned} \tag{4.3.18}$$

The result (4.3.13) is obtained by applying Lemma 4.3.3 with $\gamma = 1$, whereas $\gamma = \frac{1}{2}$ is required for (4.3.14). □

Observe that Theorem 4.3.6 allows us to estimate the error $|Z_i - \hat{Z}_i|_2$ for each i individually. This is not the case in Theorem 2.3.6 in Chapter 2 and [BD07, Theorem 10], where a time-averaged norm is used.

4.4 Monte Carlo regression scheme

Throughout this section, $(\mathbf{A}_\xi)$, $(\mathbf{A_F})$, $(\mathbf{A_H})$, $(\mathbf{A_X})$, $(\mathbf{A}'_\xi)$ and $(\mathbf{A'_F})$ are in force.

4.4.1 Algorithm

Markovian framework. Let (Y_i, Z_i) solve the discrete BSDE with data $(\xi = \Phi(X_N), f_j(X_j, y, z))$. We use the representation $(Y_i, Z_i) = (y_i(X_i), z_i(X_i))$, for deterministic, measurable, unknown functions $y_i : \mathbb{R}^d \to \mathbb{R}$ and $z_i : \mathbb{R}^d \to \mathbb{R}^q$, given in Lemma 4.2.1.

Linear least-squares regression. We denote M i.i.d. samples of a given random variable U by $U^{1:M} = \{U^1, \ldots, U^M\}$.

Given a set of *basis functions* $\{p_k : \mathbb{R}^d \to \mathbb{R} \ : \ 0 \leq k \leq K\}$, $\mathbb{R}^d$-valued random variables $X^{1:M}$ (*observations*), and $\mathbb{R}$-valued random variables $S^{1:M}$ (*response*), recall that the linear least-squares regression computes the coefficients

$$\left.\begin{array}{rcl}\mathcal{A} & = & \{\alpha \in \mathbb{R}^K \ : \ \arg\min_{\beta \in \mathbb{R}^K} \frac{1}{M} \sum_{m=1}^M |\beta \cdot p(X^m) - S^m|^2\} \\ \alpha^* & = & \arg\min_{\alpha \in \mathcal{A}} \|\alpha\|_{\mathbb{R}^K}\end{array}\right\}$$

where $p(x) = (p_1(x), \ldots, p_K(x))$ and $\|\cdot\|_{\mathbb{R}^K}$ is the usual Euclidean norm.

In implementation, we use a Singular Value Decomposition approach to compute the coefficient α^*; see [GVL96] for a detailed account.

In what follows, we will repeatedly compute linear least-squares regressions. In order to avoid repetition, we will shall say that the coefficient α^* computed by the above procedure solves

$$LS(X^{(1:M)}, S^{(1:M)}, p(x)). \tag{4.4.1}$$

Proposition 4.4.1 (Regression coefficients)**.** *Recall the least-squares regression in Definition 3.3.1. The following properties are satisfied:*

i) Linearity: if $\alpha_1 \in \mathbb{R}^K$ solves $LS\big(X^{1:M}, S_1^{1:M}, p(x)\big)$, $\alpha_2 \in \mathbb{R}^K$ solves $LS\big(X^{1:M}, S_2^{1:M}, p(x)\big)$, and $(\beta_1, \beta_2) \in \mathbb{R}^2$, then $\beta_1\alpha_1 + \beta_2\alpha_2$ solves $LS\big(X^{1:M}, \beta_1 S_1^{1:M} + \beta_2 S_2^{1:M}, p(x)\big)$.

ii) Contraction property: if $\alpha \in \mathbb{R}^K$ solves $LS\big(X^{1:M}, S^{1:M}, p(x)\big)$, then $\frac{1}{M}\sum_{m=1}^{M} |\alpha \cdot p(X^m)|^2 \leq \frac{1}{M}\sum_{m=1}^{M} |S^m|^2$.

iii) Conditional expectation solution: suppose that $\big(p(X^m)\big)_{m=1,\ldots,M}$ is measurable with respect to the σ-algebra $\mathcal{Q}$ and $\alpha \in \mathbb{R}^K$ solves $LS\big(X^{1:M}, S^{1:M}, p(x)\big)$. Then, denoting by $\mathbb{E}(S^{1:M}|\mathcal{Q})$ the random $\mathbb{R}^M$-vector with m-th entry $(\mathbb{E}(S^m|\mathcal{Q}))$, $\mathbb{E}(\alpha|\mathcal{Q})$ solves $LS\big(X^{1:M}, \mathbb{E}(S^{1:M}|\mathcal{Q}), p(x)\big)$.

The proof of Proposition 4.4.1 is exactly the same as the proof of Lemma 2.4.1 in Chapter 2. In fact, on closer inspection, the two propositions are exactly the same.

Basis functions. For each $l = 0, \ldots, q$ and $i = 0, \ldots, N-1$, we are given basis functions $(p_{l,i}^k(x))_{1\leq k\leq K_{l,i}}$. We write $p_{l,i}(x) := (p_{l,i}^1(x), \ldots, p_{l,i}^{K_{l,i}}(x))$ for the ordered vector of basis functions.

In what follows, there will be several *dynamic* least-squares regressions with observation/response pair indexed by $(i, l) \in \{0, \ldots, N-1\} \times \{0, \ldots, q\}$; the basis functions used for the (i, l)-th regression will always be $p_{l,i}(x)$, so we will omit stating explicitly which basis is being used.

Simulations. For $i = 0, \ldots, N-1$, let M_i be an integer and generate M_i copies $\{(\Omega^{(i,m)}, \mathcal{F}^{(i,m)}, \mathbb{P}^{(i,m)})$: $m = 1, \ldots, M_i\}$ of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Define $X^{i,m}$ to be the copy of X, $\Delta W^{i,m}$ to be the copy of ΔW, and $H^{i,m}$ to be the copy of the Malliavin weights H^i in $(\Omega^{(i,m)}, \mathcal{F}^{(i,m)}, \mathbb{P}^{(i,m)})$; we call these objects the *simulations*, because, in practice, one generates these objects using random number generators. Define by $(\Omega^{(M)}, \mathcal{F}^{(M)}, \mathbb{P}^{(M)})$ the product space of $(\Omega, \mathcal{F}, \mathbb{P})$ and $\bigotimes_{k,m}(\Omega^{(k,m)}, \mathcal{F}^{(k,m)}, \mathbb{P}^{(k,m)})$, and $\mathbb{E}^{(M)}$ the associated expectation operator. The clouds $\{(X^{i,1:M_i}, \Delta W^{i,1:M_i}, H^{i,1:M}$ $i = 0, \ldots, N-1\}$ are independent in $(\Omega^{(M)}, \mathcal{F}^{(M)}, \mathbb{P}^{(M)})$. By a slight abuse of notation, we shall use the notation $(\Omega, \mathcal{F}, \mathbb{P})$ to mean $(\Omega^{(M)}, \mathcal{F}^{(M)}, \mathbb{P}^{(M)})$ from now on.

We denote by $M = (M_0, \ldots, M_{N-1})$ the vector of sizes of path clouds. Such approach is very convenient for parallel computing using N processors: each processor can be in charge in simulating one path cloud and in computing the related regression coefficients.

For $i \in \{0, \ldots, N-1\}$, define the σ-algebras

$$\begin{aligned}
\mathcal{F}_i^* &:= \sigma(X_l^{k,m_k}, H_l^{k,m_k} : i < k \leq N,\ 1 \leq l \leq N,\ 1 \leq m_k \leq M_k),\\
\mathcal{F}_i^M &:= \sigma(X_i^{i,m_i} : 1 \leq m_i \leq M_i) \vee \mathcal{F}_i^*,\\
\mathcal{F}^M &:= \sigma(X_l^{k,m_k}, H_l^{k,m_k} : 0 \leq k \leq N, 1 \leq l \leq N, k < l_2 \leq N, 1 \leq m_k \leq M_k).
\end{aligned}$$

We denote by $\mathbb{E}_i^*$ ($\mathbb{P}_i^*$) the conditional expectation (probability) with respect to the σ-algebra $\mathcal{F}_i^*$, and by $\mathbb{E}_i^M$ ($\mathbb{P}_i^M$) the conditional expectation (probability) with respect to the σ-algebra $\mathcal{F}_i^M$.

Regression coefficients and approximate solutions. For every $i \in \{0 \ldots, N-1\}$ and $l \in \{1, \ldots, q\}$, define the random variables

$$\Psi(i,m) = \Phi(X_N^{i,m}) + \sum_{j=i+1}^{N-1} f_j(X_j^{i,m}, y_j^M(X_j^{i,m}), z_j^M(X_j^{i,m}))\Delta_j, \tag{4.4.2}$$

$$\Xi_l(i,m) = \Phi(X_N^{i,M})H_{l,N}^{i,m} + \sum_{j=i+1}^{N-1} f_j(X_j^{i,m}, y_j^M(X_j^{i,m}), z_j^M(X_j^{i,m}))H_{l,j}^{i,m}\Delta_j. \tag{4.4.3}$$

Recall the definition of $LS(\cdot)$ in (4.4.1). Let $\alpha^M_{0,i}$ be the random vector in $\mathbb{R}^{K_{0,i}}$ solving

$$LS\big(X_i^{i,1:M_i}, \Psi(i, 1 : M_i), p_{0,i}(x)\big).$$

For $l = 1, \ldots, q$, let $\alpha^M_{l,i}$ be the random $\mathbb{R}^{K_{l,i}}$-vector solving $LS\big(X_i^{i,1:M_i}, \Xi(i, l, 1 : M_i)p_{l,i}(x)\big)$. By Corollary 4.3.5, the solutions y_i and z_i of the discrete BSDE are bounded by $C_{y,i}$ and $C_{z,i}$, respectively. We use these soft thresholds for the least-squares regression approximations: for $(y, z) \in \mathbb{R} \times \mathbb{R}^q$, define

$$[y]_{y,i} := -C_{y,i} \vee y \wedge C_{y,i}, \quad [z_l]_{z,i} := -C_{z,i} \vee z_l \wedge C_{z,i},$$

and set

$$y^M_i(x) := [\alpha^M_{0,i} \cdot p_{0,i}(x)]_{y,i}, \quad z^M_{l,i}(x) := [\alpha^M_{l,i} \cdot p_{l,i}(x)]_{z,i}. \tag{4.4.4}$$

This restricts the range of y^M_i and z^M_i and prevents numerical instability due to overfitting.

Additional coefficients. Recall the definition of $LS(\cdot)$ in (4.4.1). In the following error analysis, we make use of the regression coefficients $\beta^M_{l,i}$ for $i = 0, \ldots, N-1$ and $l = 0, \ldots, q$, which are the random vectors in $\mathbb{R}^{K_{l,i}}$ solving $LS\big(X_i^{i,1:M_i}, \tilde{\Psi}(i, 1 : M_i), p_{0,i}(x)\big)$ for $l = 0$ and $LS\big(X_i^{i,1:M_i}, \tilde{\Xi}_l(i, 1 : M_i), p_{l,i}(x)\big)$ for $l > 0$, where

$$\tilde{\Psi}(i, m) = \Phi(X_N^{i,m}) + \sum_{j=i+1}^{N-1} f_j(X_j^{i,m}, y_j(X_j^{i,m}), z_j(X_j^{i,m}))\Delta_j,$$

$$\tilde{\Xi}_l(i, m) = \Phi(X_N^{i,m})H^{i,m}_{l,N} + \sum_{j=i+1}^{N-1} f_j(X_j^{i,m}, y_j(X_j^{i,m}), z_j(X_j^{i,m}))H^{i,m}_{l,j}\Delta_j.$$

Intermediate processes. Let $(X_k^{i,x})_{k\geq i}$ be the Markov chain given by $X_k^{i,x} = V_k^i(x)$. and $(H_j^{i,x})_{j>i}$ be the Malliavin weights given by $h^i_j(x)$; see $(\mathbf{A_X})$ and $(\mathbf{A'_H})$. Define the following functions:

$$\left.\begin{array}{l} \bar{y}^M_i(x) := \int \Psi(i, x_{i+1}, \ldots, x_N)\mathrm{d}\mu^x_i(x_{i+1}, \ldots, x_N), \\ \bar{z}^M_i(x) := \int \Xi(i, x_{i+1}, \ldots, x_N, h_{i+1}, \ldots, h_N)\mathrm{d}\lambda^x_i(x_{i+1}, \ldots, x_N, h_{i+1}, \ldots, h_N) \end{array}\right\} \tag{4.4.5}$$

where μ^x_i is the law of $(X^{i,x}_{i+1}, \ldots, X^{i,x}_N)$, λ^x_i is the law of $(X^{i,x}_{i+1}, \ldots, X^{i,x}_N, H^{i,x}_{i+1}, \ldots, H^{i,x}_N)$, and

$$\Psi(i, x_{i+1}, \ldots, x_N) = \Phi(x_N) + \sum_{j=i+1}^{N-1} f_j(x_j, y^M_j(x_j), z^M_j(x_j))\Delta_j, \tag{4.4.6}$$

$$\begin{aligned} \Xi(i, x_{i+1}, \ldots, x_N, &h_{i+1}, \ldots, h_N) \\ &= \Phi(x_N)h_N + \sum_{j=i+1}^{N-1} f_j(x_j, y^M_j(x_j), z^M_j(x_j))h_j\Delta_j. \end{aligned} \tag{4.4.7}$$

Lemma 4.4.2. *Assume* $(\mathbf{A_X})$, $(\mathbf{A'_\xi})$*and* $(\mathbf{A'_F})$. *The following representation holds for all* m *and* *i:*

$$\bar{y}_i^M(X_i^{i,m}) = \mathbb{E}_i^M[\Psi(i, X_{i+1}^{i,m}, \dots, X_N^{i,m})],$$
$$\bar{z}_i^M(X_i^m) = \mathbb{E}_i^M[\Xi(i, X_{i+1}^{i,m}, \dots, X_N^{i,m}, H_{i+1}^{i,m}, \dots, H_N^{i,m})].$$

The proof of Lemma 4.4.2 is analogous to the proof of Lemma 4.2.1.

Lemma 4.4.3. *There exist explicit time-dependent constants $C_{\bar{y},i}, C_{\bar{z},i}$ such that, for all i, $|\bar{y}_i^M(x)| \leq C_{\bar{y},i}$ and $|\bar{z}_i^M(x)| \leq C_{\bar{z},i}$, $\mathbb{P} \circ X_i^{-1}$-a.s.*

Proof. We give details for $\bar{z}_i^M(X_i^m)$: the method will be used again later. We upper bound the conditional second moment of $\Xi(i, X_{i+1}^{i,m}, \dots, X_N^{i,m}, H_{i+1}^{i,m}, \dots, H_N^{i,m})$:

$$\begin{aligned}
&\big(\mathbb{E}_i^M[\Xi(i, X_{i+1}^{i,m}, \dots, X_N^{i,m}, H_{i+1}^{i,m}, \dots, H_N^{i,m})]|^2\big)^{\frac{1}{2}} \\
&\leq \big(|\mathbb{E}_i^M[(\Phi(X_N^{i,m}) - \mathbb{E}_i^M[\Phi(X_N^m)])H_N^{i,m}]|^2\big)^{\frac{1}{2}} \\
&\quad + \sum_{j=i+1}^{N-1} \big(|\mathbb{E}_i^M[f_j(X_j^{i,m}, y_j^M(X_j^{i,m}), z_j^M(X_j^{i,m}))H_j^{i,m}]|^2\big)^{\frac{1}{2}} \Delta_j \\
&\leq C_M \frac{\big(\mathbb{V}\mathrm{ar}(\xi^{i,m}|X_i^{i,m})\big)^{\frac{1}{2}}}{\sqrt{T-t_i}} + \sum_{j=i+1}^{N-1} \frac{C_M}{\sqrt{t_j - t_i}} \Big(\frac{C_f}{(T-t_j)^{1-\theta_c}} + L_f \frac{C_{y,j} + C_{z,j}\sqrt{q}}{(T-t_j)^{(1-\theta)/2}}\Big) \Delta_j
\end{aligned}$$

Applying (4.3.10) and Lemma 4.3.1, we derive that the term above is a.s bounded by

$$C_M \Big(\operatorname*{ess\,sup}_{\omega} \frac{(\mathbb{V}\mathrm{ar}(\xi|X_i))^{\frac{1}{2}}}{\sqrt{T-t_i}} + \frac{C_f B_{\frac{1}{2},\theta_c}}{(T-t_i)^{\frac{1}{2}-\theta_c}} + L_f B_{\frac{1}{2},\frac{\theta}{2}} C_{y,z} \frac{(\sqrt{T} + \sqrt{q})}{(T-t_i)^{(1-\theta)/2}} \Big) =: C_{\bar{z},i}.$$

Similarly,

$$\begin{aligned}
&\big(\mathbb{E}_i^M[\Psi(i, X_{i+1}^{i,m}, \dots, X_N^{i,m})]^2\big)^{\frac{1}{2}} \\
\leq & C_{\bar{y},i} := \big(C_\xi + C_f B_{1,\theta_c}(T-t_i)^{\theta_c} + L_f B_{1,\frac{\theta}{2}} C_{y,z}(\sqrt{T} + \sqrt{q})(T-t_i)^{\frac{\theta}{2}}\big).
\end{aligned}$$

□

Norms. For a function $f : \mathbb{R}^d \to \mathbb{R}^r$, define the square norms

$$\|f\|_{i,M}^2 := \frac{1}{M_i} \sum_{m=1}^{M_i} |f(X_i^{i,m})|^2, \qquad \|f\|_{i,\infty}^2 := \int_{\mathbb{R}^d} |f(x)|^2 \mathrm{d}\mathbb{P} \circ X_i^{-1}(x).$$

The first one is the empirical approximation of the second one.

4.4.2 Error analysis

Local error terms. We shall make use of the following error terms:

$$T_{1,i}^{Y,M} := \mathbb{E}[\inf_{\alpha \in \mathbb{R}^{K_{0,i}}} \|\alpha \cdot p_{0,i} - y_i\|_{i,M}^2], \tag{4.4.8}$$

$$T_{1,i}^{Z,M} := \sum_{l=1}^{q} \mathbb{E}[\inf_{\alpha \in \mathbb{R}^{K_{l,i}}} \|\alpha \cdot p_{l,i} - z_{l,i}\|_{i,M}^2], \tag{4.4.9}$$

$$T_{2,i}^{Y,M} := \mathbb{E}[\|(\alpha_{0,i}^M - \mathbb{E}_i^M[\alpha_{0,i}^M]) \cdot p_{0,i}\|_{i,M}^2], \tag{4.4.10}$$

$$T_{2,i}^{Z,M} := \sum_{l=1}^{q} \mathbb{E}[\|(\alpha_{l,i}^M - \mathbb{E}_i^M[\alpha_{l,i}^M]) \cdot p_{l,i}\|_{i,M}^2]. \tag{4.4.11}$$

For $i = 0, \ldots, N-1$, let $\varepsilon_{i,A}^Y, \varepsilon_{i,A}^Z, \varepsilon_{i,B}^Y, \varepsilon_{i,B}^Z > 0$. We need the following error events:

$$A_i^{Y,M} := \{\varepsilon_{i,A}^Y + 2\|\bar{y}_i^M - y_i\|_{i,\infty}^2 < \|\bar{y}_i^M - y_i\|_{i,M}^2\}, \tag{4.4.12}$$

$$A_i^{Z,M} := \{\exists l \in \{1, \ldots, q\} \; s.t. \; \varepsilon_{i,A}^Z + 2\|\bar{z}_{l,i}^M - z_{l,i}\|_{i,\infty}^2 < \|\bar{z}_{l,i}^M - z_{l,i}\|_{i,M}^2\}, \tag{4.4.13}$$

$$B_i^{Y,M} := \{\varepsilon_{i,B}^Y + 2\|y_i^M - y_i\|_{i,M}^2 < \|y_i^M - y_i\|_{i,\infty}^2\}, \tag{4.4.14}$$

$$B_i^{Z,M} := \{\exists l \in \{1, \ldots, q\} \; s.t. \; \varepsilon_{i,B}^Z + 2\|z_{l,i}^M - z_{l,i}\|_{i,M}^2 < \|z_{l,i}^M - z_{l,i}\|_{i,\infty}^2\}, \tag{4.4.15}$$

The aim is to determine a rate of convergence for $\mathbb{E}\|y_i - y_i^M\|_{i,M}$ and $\mathbb{E}\|z_i - z_i^M\|_{i,M}$ using the local error terms.

Theorem 4.4.4. *For $0 \le i \le j \le N-1$, define*

$$\left.\begin{array}{rl} \frac{1}{3}\mathcal{E}_{Y,i}^M & := T_{1,i}^{Y,M} + T_{2,i}^{Y,M} + (C_{y,i} + C_{\bar{y},i})^2 \mathbb{P}(A_i^{Y,M}) + \varepsilon_{i,A}^Y, \\ \frac{1}{3}\mathcal{E}_{Z,i}^M & := T_{1,i}^{Z,M} + T_{2,i}^{Z,M} + (C_{z,i} + C_{\bar{z},i})^2 \mathbb{P}(A_i^{Z,M}) + \varepsilon_{i,A}^Z, \\ \mathcal{E}_{1,i}^M & := (\mathcal{E}_{Y,i}^M)^{\frac{1}{2}} + (\mathcal{E}_{Z,i}^M)^{\frac{1}{2}}, \\ \frac{1}{2}\mathcal{E}_{2,i}^M & := 4C_{y,i}^2 \mathbb{P}(B_i^{Y,M}) + 4C_{z,i}^2 \mathbb{P}(B_i^{Z,M}) + \varepsilon_{i,B}^Y + \varepsilon_{i,B}^Z. \end{array}\right\} \tag{4.4.16}$$

There exists a constant $C \ge 0$ that depends only on $C_y^{(2)}$, $C_z^{(2)}$, T, L_f and θ such that

$$\mathbb{E}[\|y_i - y_i^M\|_{i,M}^2]^{1/2} \le (\mathcal{E}_{Y,i}^M)^{\frac{1}{2}} + C \sum_{j=i+1}^{N-1} \frac{\{\mathcal{E}_{1,j}^M + (\mathcal{E}_{2,j}^M)^{\frac{1}{2}}\}\Delta_j}{(T - t_j)^{(1-\theta)/2}}, \tag{4.4.17}$$

$$\mathbb{E}[\|z_i - z_i^M\|_{i,M}^2]^{1/2} \le (\mathcal{E}_{Z,i}^M)^{\frac{1}{2}} + C \sum_{j=i+1}^{N-1} \frac{\{\mathcal{E}_{1,j}^M + (\mathcal{E}_{2,j}^M)^{\frac{1}{2}}\}\Delta_j}{(T - t_j)^{(1-\theta)/2}\sqrt{t_j - t_i}}. \tag{4.4.18}$$

Proof. Define the intermediate error terms

$$\left.\begin{array}{rl} \frac{1}{3}G_{Y,i}^M & := \inf_{\alpha \in \mathbb{R}^{K_{0,i}}} \|\alpha \cdot p_{0,i} - y_i\|_{i,M}^2 + \|(\alpha_{0,i}^M - \mathbb{E}_i^M[\alpha_{0,i}^M]) \cdot p_{0,i}\|_{i,M}^2 \\ & \quad +(C_{y,i} + C_{\bar{y},i})^2 \mathbf{1}_{A_i^{Y,M}} + \varepsilon_{i,A}^Y, \\ \frac{1}{3}G_{Z,i}^M & := \sum_{l=1}^{q} \left\{ \inf_{\alpha \in \mathbb{R}^{K_{l,i}}} \|\alpha \cdot p_{l,i} - z_{l,i}\|_{i,M}^2 + \|(\alpha_{l,i}^M - \mathbb{E}_i^M[\alpha_{l,i}^M]) \cdot p_{l,i}\|_{i,M}^2 \right\} \\ & \quad +(C_{z,i} + C_{\bar{z},i})^2 \mathbf{1}_{A_i^{Z,M}} + \varepsilon_{i,A}^Z, \\ \frac{1}{2}G_{2,i}^M & := 4C_{y,i}^2 \mathbf{1}_{B_i^{Y,M}} + 4C_{z,i}^2 \mathbf{1}_{B_i^{Z,M}} + \varepsilon_{i,B}^Y + \varepsilon_{i,B}^Z \end{array}\right\} \tag{4.4.19}$$

and observe that $\mathbb{E}[G_{u,i}^M] = \mathcal{E}_{u,i}^M$ for the indexes $u \in \{Y, Z, 2\}$. Define also $G_{1,i}^M := (G_{Y,i}^M)^{\frac{1}{2}} + (G_{Z,i}^M)^{\frac{1}{2}}$. Using the triangle inequality, it follows that

$$\begin{aligned}\|y_i - y_i^M\|_{i,M} \leq \|y_i - \mathbb{E}_i^M[\beta_{0,i}^M] \cdot p_{0,i}\|_{i,M} + \|\mathbb{E}_i^M[\beta_{0,i}^M - \alpha_{0,i}^M] \cdot p_{0,i}\|_{i,M} \\ + \|(\mathbb{E}_i^M[\alpha_{0,i}^M] - \alpha_{0,i}^M) \cdot p_{0,i}\|_{i,M}\end{aligned} \tag{4.4.20}$$

Recall the definition of $LS(\cdot)$ in (4.4.1). Now, since $\mathbb{E}_i^M[\tilde{\Psi}(i,m)] = y_i(X_i^{i,m})$, Proposition 4.4.1(iii) implies that $\mathbb{E}_i^M[\beta_{0,i}^M]$ is the random vector in $\mathbb{R}^{K_{0,i}}$ solving $LS\big(X_i^{i,1:M_i}, y_i(X_i^{i,1:M_i}), p_{0,i}(x)\big)$. From Lemma 4.4.2 we have that $\mathbb{E}_i^M[\Psi(i, X_{i+1}^{i,m}, \ldots, X_N^{i,m})] = \bar{y}_i(X_i^{i,m})$, therefore Proposition 4.4.1(iii) implies that $\mathbb{E}_i^M[\alpha_{0,i}^M]$ is the random vector in $\mathbb{R}^{K_{0,i}}$ solving $LS\big(X_i^{i,1:M_i}, \bar{y}_i^M(X_i^{i,1:M_i}), p_{0,i}(x)\big)$. By the linearity property Proposition 4.4.1(i), we have that $\mathbb{E}_N^M[\beta_{0,i}^M - \alpha_{0,i}^M]$ is random vector in $\mathbb{R}^{K_{0,i}}$ solving $LS\big(X_i^{i,1:M_i}, (y_i - \bar{y}_i^M)(X_i^{i,1:M_i}), p_{0,i}(x)\big)$. Thus, using additionally the contraction property Proposition 4.4.1(ii), and Jensen's inequality, it follows that

$$\|\mathbb{E}_i^M[\beta_{0,i}^M - \alpha_{0,i}^M] \cdot p_{0,i}\|_{i,M}^2 \leq \|y_i - \bar{y}_i\|_{i,M}^2 \leq \varepsilon_{i,A}^Y + 2\|y_i - \bar{y}_i\|_{i,\infty}^2 + (C_{y,i} + C_{\bar{y},i})^2 \mathbf{1}_{A_i^{Y,M}} \tag{4.4.21}$$

Substituting (4.4.21) into (4.4.20) yields

$$\|y_i - y_i^M\|_{i,M} \leq (G_{Y,i}^M)^{\frac{1}{2}} + \sqrt{2}\|y_i - \bar{y}_i\|_{i,\infty}. \tag{4.4.22}$$

By the same arguments, we obtain

$$\|z_i - z_i^M\|_{i,M} \leq (G_{Z,i}^M)^{\frac{1}{2}} + \sqrt{2}\|z_i - \bar{z}_i^M\|_{i,\infty}. \tag{4.4.23}$$

Define for $j = 0, \ldots, N-1$

$$\Theta_j := \|y_j - y_j^M\|_{j,M} + \|z_j - z_j^M\|_{j,M}. \tag{4.4.24}$$

Now, $(\bar{y}_i^M(X_i), \bar{z}_i^M(X_i))_{i \geq 0}$ solves the discrete BSDE $(\xi, \bar{f}_j(y,z) = f_j(X_j, y_j^M(X_j), z_j^M(X_j)))$ with respect to the filtration $\{\mathcal{F}^M \vee \mathcal{F}_i\}_{0 \leq i \leq N}$. Now, since $\bar{y}_i(x)$ and $\bar{z}_i(x)$ are $\mathcal{F}^M \otimes \mathcal{B}(\mathbb{R}^d)$-measurable, it follows that

$$\|y_i - \bar{y}_i\|_{i,\infty}^2 = \mathbb{E}[|y_i(X_i) - \bar{y}_i(X_i)|^2 | \mathcal{F}^M] \quad \text{and} \quad \|z_i - \bar{z}_i\|_{i,\infty}^2 = \mathbb{E}[|z_i(X_i) - \bar{z}_i(X_i)|^2 | \mathcal{F}^M]$$

thanks to Lemma 4.2.2. Using Proposition 4.3.4, (4.4.22) and the Lipschitz continuity of f, we obtain

$$\begin{aligned}\|y_i - y_i^M\|_{i,M} &\leq (G_{Y,i}^M)^{\frac{1}{2}} + \sqrt{2} C_y^{(2)} L_f \sum_{j=i+1}^{N-1} \frac{\{\|y_j - y_j^M\|_{j,\infty} + \|z_j - z_j^M\|_{j,\infty}\}\Delta_j}{(T - t_j)^{(1-\theta)/2}} \\ &\leq (G_{Y,i}^M)^{\frac{1}{2}} + \sqrt{2} C_y^{(2)} L_f \sum_{j=i+1}^{N-1} \frac{\{(G_{2,j}^M)^{\frac{1}{2}} + \sqrt{2}\Theta_j\}\Delta_j}{(T - t_j)^{(1-\theta)/2}}\end{aligned} \tag{4.4.25}$$

where we have used the definitions (4.4.14) and (4.4.15), and the a.s. bounds on y_i and z_i from

Corollary 4.3.5. Likewise, from (4.4.23) we obtain:

$$\|z_i - z_i^M\|_{i,M} \le (G_{Z,i}^M)^{\frac{1}{2}} + \sqrt{2}C_z^{(2)} L_f \sum_{j=i+1}^{N-1} \frac{\{(G_{2,j}^M)^{\frac{1}{2}} + \sqrt{2}\Theta_j\}\Delta_j}{(T-t_j)^{(1-\theta)/2}\sqrt{t_j - t_i}}. \tag{4.4.26}$$

Combining (4.4.25) and (4.4.26), we obtain

$$\Theta_i \le G_{1,i}^M + \sqrt{2}(C_z^{(2)} + \sqrt{T}C_y^{(2)})L_f \sum_{j=i+1}^{N-1} \frac{(G_{2,j}^M)^{\frac{1}{2}} + \sqrt{2}\Theta_j}{(T-t_j)^{(1-\theta)/2}\sqrt{t_j - t_i}}\Delta_j.$$

Using Lemma 4.3.2 in the setting $u_j = \Theta_j$,

$$w_j = G_{1,j}^M + \sqrt{2}(C_z^{(2)} + \sqrt{T}C_y^{(2)})L_f \sum_{k=j+1}^{N-1} \frac{(G_{2,k}^M)^{\frac{1}{2}}\Delta_k}{(T-t_k)^{(1-\theta)/2}\sqrt{t_k - t_j}},$$

$C_u = 2(C_z^{(2)} + \sqrt{T}C_y^{(2)})L_f$, $\alpha = 0$ and $\beta = \frac{\theta}{2}$, we obtain

$$\Theta_i \le CG_{1,i}^M + C\sum_{j=i+1}^{N-1} \frac{\{G_{1,j}^M + (G_{2,j}^M)^{\frac{1}{2}}\}\Delta_j}{(T-t_j)^{(1-\theta)/2}\sqrt{t_j - t_i}} + C\sum_{j=i+1}^{N-1} \frac{\Theta_j\Delta_j}{(T-t_j)^{(1-\theta)/2}} \tag{4.4.27}$$

for a constant C that depends only on $C_y^{(2)}$, $C_z^{(2)}$, T, L_f and θ (to simplify the sum of $G_{2,j}^M$-terms, we have used the general inequality (4.5.4) with $\delta = 0$ and $\gamma = \frac{1}{2}$). Using Lemma 4.3.3, combined with (4.4.27), to upper bound the sums in (4.4.25) and (4.4.26) - we take $\gamma = 1$ for (4.4.25), and with $\gamma = \frac{1}{2}$ for (4.4.26) - it follows that

$$\|y_i - y_i^M\|_{i,M} \le (G_{Y,i}^M)^{\frac{1}{2}} + C\sum_{j=i+1}^{N-1} \frac{\{G_{1,j}^M + (G_{2,j}^M)^{\frac{1}{2}}\}\Delta_j}{(T-t_j)^{(1-\theta)/2}}, \tag{4.4.28}$$

$$\|z_i - z_i^M\|_{i,M} \le (G_{Z,i}^M)^{\frac{1}{2}} + C\sum_{j=i+1}^{N-1} \frac{\{G_{1,j}^M + (G_{2,j}^M)^{\frac{1}{2}}\}\Delta_j}{(T-t_j)^{(1-\theta)/2}\sqrt{t_j - t_i}}. \tag{4.4.29}$$

for a constant C that depends only on $C_y^{(2)}$, $C_z^{(2)}$, T, L_f and θ. We complete the proof by squaring and taking expectations in (4.4.28) and (4.4.29), and then using Minkowski's inequality. □

Theorem 4.4.5. *For any* $i \in \{0, \dots, N-1\}$,

$$T_{1,i}^{Y,M} \le \inf_{\alpha \in \mathbb{R}^{K_{0,i}}} \mathbb{E}[|y_i(X_i) - \alpha \cdot p_{0,i}(X_i)|^2], \qquad T_{2,i}^{Y,M} \le \frac{C_{\bar{y},i}^2 K_{0,i}}{M_i}, \tag{4.4.30}$$

$$T_{1,i}^{Y,M} \le \sum_{l=0}^{q} \inf_{\alpha \in \mathbb{R}^{K_{l,i}}} \mathbb{E}[|z_{l,i}(X_i) - \alpha \cdot p_{l,i}(X_i)|^2], \qquad T_{2,i}^{Z,M} \le \sum_{l=1}^{q} \frac{C_{\bar{z},i}^2 K_{l,i}}{M_i}. \tag{4.4.31}$$

Moreover,

$$\mathbb{P}_i^M(A_i^{Y,M}) \le \exp(-\frac{M_i\theta^4(\varepsilon_{i,A}^Y)^2(T-t_i)^{2\theta}}{224L_f^4(q^2+T^2)C_{y,z}^4}), \quad \mathbb{P}_i^M(A_i^{Z,M}) \le \exp(-\frac{M_i(\varepsilon_{i,A}^Z)^2(T-t_i)^{2(1-\theta)}}{8C_M^4B_{\frac{\theta}{2},\frac{1}{2}}^4L_f^4(q^2+T^2)C_{y,z}^4}). \tag{4.4.32}$$

and, for $0 \le \varepsilon_{i,B}^Y \le 24C_{y,i}$ *and* $0 \le \varepsilon_{i,B}^Z \le 24C_{z,i}$,

$$\mathbb{P}(B_i^{Y,M}) \le 12\Big(\frac{1056C_{y,i}^2}{5\varepsilon_{B,i}^Y}\Big)^{2(K_{0,i}+1)}\exp\Big(-\frac{M_i\varepsilon_{B,i}^Y}{507C_{y,i}^2}\Big), \tag{4.4.33}$$

$$\mathbb{P}(B_i^{Z,M}) \le 12\sum_{l=1}^q\Big(\frac{1056C_{z,i}^2}{5\varepsilon_{B,i}^Z}\Big)^{2(K_{l,i}+1)}\exp\Big(-\frac{M_i\varepsilon_{B,i}^Z}{507C_{z,i}^2}\Big). \tag{4.4.34}$$

Proof. Let us consider the value $|y_i(x)-\bar{y}_i(x)|$. Thanks to definition (4.4.5), (4.2.1), and Minkowski's inequality, and the uniform bound (4.3.10) it follows that

$$\begin{aligned}
&|y_i(x)-\bar{y}_i(x)| \\
&\le \sum_{j=i+1}^{N-1}\int|f_j(x_j,y_j(x_j),z_j(x_j))-f_j(x_j,y_j^M(x_j),z_j^M(x_j))|d\mu_i^x\Delta_j \\
&\le \sum_{j=i+1}^{N-1}\frac{L_f\Delta_j}{(T-t_j)^{(1-\theta)/2}}\int\{|y_j(x_j)-y_j^M(x_j)|+|z_j(x_j)-z_j^M(x_j)|\}d\mu_i^x \\
&\le \sum_{j+1}^{N-1}\frac{L_f\Delta_j\{2C_{y,j}+2\sqrt{q}C_{z,j}\}}{(T-t_j)^{(1-\theta)/2}} \le \frac{4}{\theta}L_fC_{y,z}(\sqrt{q}+\sqrt{T})(T-t_i)^{\frac{\theta}{2}}
\end{aligned} \tag{4.4.35}$$

The functions $|y_i(\cdot)-\bar{y}_i(\cdot)|$ are $\mathcal{F}_i^M$- measurable, therefore the random variables $\{|y_i(X_i^{i,m})-\bar{y}_i(X_i^{i,m})| : 1 \le m \le M_i\}$ are $\mathcal{F}_i^M$-conditionally independent and bounded. We bound $\mathbb{P}_i^M(A_{Y,i}^M)$ using the conditional Hoeffding's inequality Lemma 4.5.1:

$$\mathbb{P}_i^M(A_i^{Y,M}) \le \exp(-\frac{M_i\theta^4(\varepsilon_{i,A}^Y)^2}{224L_f^4(q^2+T^2)C_{y,z}^4(T-t_i)^{2\theta}}). \tag{4.4.36}$$

We can also bound $|z_i(x)-\bar{z}_i(x)|$ by applying (4.4.5), (4.2.1), Minkowski's inequality, the Cauchy-Schwarz inequality, the uniform bound (4.3.10) and Lemma 4.3.1:

$$\begin{aligned}
&|z_i(x)-\bar{z}_i(x)| \\
&\le \sum_{j=i+1}^{N-1}\int|\big(f_j(x_j,y_j(x_j),z_j(x_j))-f_j(x_j,y_j^M(x_j),z_j^M(x_j))\big)h_j|d\lambda_i^x\Delta_j \\
&\le \sum_{j=i+1}^{N-1}\frac{C_M(\int|f_j(x_j,y_j(x_j),z_j(x_j))-f_j(x_j,y_j^M(x_j),z_j^M(x_j))|^2d\mu_i^x)^{\frac{1}{2}}\Delta_j}{\sqrt{t_j-t_i}} \\
&\le \sum_{j+1}^{N-1}\frac{C_ML_f\Delta_j\{2C_{y,j}+2\sqrt{q}C_{z,j}\}}{(T-t_j)^{(1-\theta)/2}\sqrt{t_j-t_i}} \le \frac{2C_MB_{\frac{\theta}{2},\frac{1}{2}}L_fC_{y,z}(\sqrt{q}+\sqrt{T})}{(T-t_i)^{(1-\theta)/2}}
\end{aligned} \tag{4.4.37}$$

The functions $|z_i(\cdot) - \bar{z}_i(\cdot)|$ are $\mathcal{F}_i^M$- measurable, therefore the random variables $\{|z_i(X_i^{i,m}) - \bar{z}_i(X_i^{i,m})| : 1 \leq m \leq M_i\}$ are $\mathcal{F}_i^M$-conditionally independent and bounded. We bound $\mathbb{P}_i^M(A_i^{Z,M})$ using the above bound and the conditional Hoeffding's inequality:

$$\mathbb{P}_i^M(A_i^{Z,M}) \leq \exp(-\frac{M_i(\varepsilon_{i,A}^Z)^2(T-t_i)^{2(1-\theta)}}{8C_M^4 B_{\frac{\theta}{2},\frac{1}{2}}^4 L_f^4(q^2+T^2)C_{y,z}^4}).$$

Recall the bounds on $|y_i(x) - y_i^M(x)| \leq 2C_{y,i}$ (resp. $|z_i(x) - z_i^M(x)| \leq 2C_{z,i}$). To bound $\mathbb{P}_i^M(B_i^{Y,M})$ (resp. $\mathbb{P}_i^M(B_i^{Z,M})$), we proceed exactly as in the proof of the bounds on $\mathbb{P}(C_k^{Y,M})$ in Theorem 2.4.5; we must only update the bounds on the functions. □

One should note that the bound on $\mathbb{P}_i^M(A_i^{Z,M})$ in (4.4.32) is obtained using the uniform bound (4.3.10). Assuming extra smoothness of the terminal function $x \mapsto \Phi(x)$, as in the remarks at the end of Corollary 4.3.5, would improve the estimates in (4.4.35) and (4.4.37), whence also the bound on $\mathbb{P}_i^M(A_i^{Z,M})$. Additionally, one can make more precise statements on the bounds of $\mathbb{P}_i^M(B_i^{Z,M})$ in (4.4.32) and $T_{2,i}^{Z,M}$ in (4.4.31); we summarize this in the following lemma.

Lemma 4.4.6. *Assume that the terminal condition $x \mapsto \Phi(x)$ is θ_Φ-Hölder continuous and $\mathbb{E}_i[|X_N - X_i|^2] \leq C_X(T - t_i)$. Then*

$$\begin{aligned}
\mathbb{P}_i^M(A_i^{Z,M}) &\leq \exp(-CM_i(\varepsilon_{i,A}^Z)^2(T-t_i)^{2(1-\theta_c\wedge\frac{\theta_\Phi}{2}-\theta)}), \\
\mathbb{P}_i^M(B_i^{Z,M}) &\leq 12\sum_{l=1}^q \exp\Big(-C\ln\big(\varepsilon_{B,i}^Z(T-t_i)^{1-\theta_c\wedge\frac{\theta_\phi}{2}}\big)(K_{l,i}+1) - CM_i\varepsilon_{B,i}^Z(T-t_i)^{1-\theta_c\wedge\frac{\theta_\Phi}{2}}\Big), \\
T_{2,i}^{Z,M} &\leq C\sum_{l=1}^q \frac{K_{l,i}}{M_i(T-t_i)^{1-\theta_c\wedge\frac{\theta_\Phi}{2}}},
\end{aligned}$$

for a constant C depending on $T, C_M, L_f, C_f, \theta, \theta_c$, C_X, C_ξ, q, and the Hölder coefficient of Φ only.

Proof. Since Φ is Hölder continuous and bounded, $|Z_i| \leq C(T-t_i)^{(\theta_c\wedge\frac{\theta_\Phi}{2}-1)/2}$. The proof is then completed in the same way as Theorem 4.4.5. □

4.4.3 Complexity analysis

We assume that $x \mapsto \Phi(x)$ is θ_Φ-Hölder continuous. In this section, C depends on $T, C_M, L_f, C_f, \theta, \theta_c$, C_X, C_ξ, q, and the Hölder coefficient of Φ only, but it's value may change from line to line.

Typically, the discretization error of the BSDE (4.1.2), i.e.

$$\mathcal{E}(\pi) := \max_{0\leq i\leq N-1}\sup_{t_i\leq t\leq t_{i+1}} \mathbb{E}[|Y_t - Y_{t_i}|^2] + \sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} \mathbb{E}[|Z_t - Z_{t_i}|^2]dt$$

can be bounded above by $\mathcal{E}(\pi) \leq CN^{-2\theta_{\text{conv}}}$ when the terminal value and driver satisfy suitable conditions; this was the topic of Chapter 1. In this section, we will be concerned with setting those parameters in a way so that

$$\mathcal{E}(M) := \max_{0\le i\le N-1} \mathbb{E}[\|y_i - y_i^M\|_{i,M}^2] + \sum_{i=0}^{N-1} \mathbb{E}[\|z_i - z_i^M\|_{i,M}^2]\Delta_i \le CN^{-2\theta_{\text{conv}}} \tag{4.4.38}$$

From Theorems 4.4.4 and 4.4.5, we have a systematic way to control the error terms $\|y_i - y_i^M\|_{i,M}$ and $\|z_i - z_i^M\|_{i,M}$ by judicious choice of the basis functions $p_{l,i}(x)$ and the size of the sample clouds M_i: in order to ensure that $\mathcal{E}(M) \le CN^{-2\theta_{\text{conv}}}$, it is sufficient to choose basis functions $p_{l,i}$ and select simulation sizes $(M_0, \ldots, M_{N-1})$ so that the bias terms $T_{1,i}^{\cdot,M}$, the variance terms $T_{2,i}^{\cdot,M}$, and the conditional probabilities of the large deviation events $A_i^{\cdot,M}$ and $B_i^{\cdot,M}$ by $CN^{-2\theta_{\text{conv}}}$ to ensure that $\mathcal{E}(M) \le CN^{-2\theta_{\text{conv}}}$.

We briefly remark that bounding the error $\mathcal{E}(M)$ is equivalent to bounding the error when the norm $\|\cdot\|_{i,M}$ is replaced with $\|\cdot\|_{i,\infty}$; this norm is equivalent to that used in $\mathcal{E}(\pi)$. Indeed, using the same terminology as in (4.4.16),

$$\max_{0\le i\le N-1} \mathbb{E}[\|y_i - y_i^M\|_{i,\infty}^2] + \sum_{i=0}^{N-1} \mathbb{E}[\|z_i - z_i^M\|_{i,\infty}^2]\Delta_i \le C\sum_{i=0}^{N-1} \mathcal{E}_{2,i}^M \Delta_i + \mathcal{E}(M)$$

and, as we shall see below, in order to set $\mathcal{E}(M) \le CN^{-2\theta_{\text{conv}}}$, we will set $\mathcal{E}_{2,i}^M \le CN^{-2\theta_{\text{conv}}}$ anyway.

Bias terms. First, we refine the estimates for $T_{1,i}^{\cdot,M}$. Using Theorem 4.4.5 and the solution (Y_t, Z_t) of the continuous time BSDE (4.1.2), we obtain

$$\begin{aligned} T_{1,i}^{Y,M} &\le \inf_{\alpha\in\mathbb{R}^{K_{0,i}}} \mathbb{E}[|y_i(X_i) - \alpha\cdot p_{0,i}(X_i)|^2] \\ &\le C \inf_{\alpha\in\mathbb{R}^{K_{0,i}}} \mathbb{E}[|Y_{t_i} - \alpha\cdot p_{0,i}(X_{t_i})|^2] + C\mathbb{E}[|Y_{t_i} - Y_i|^2] \end{aligned}$$

Defining Z_{t_i} by the representation formula 4.1.3, a similar expression for $T_{1,i}^{Z,M}$ is available

$$T_{1,i}^{Y,M} \le C\sum_{l=1}^{q} \inf_{\alpha\in\mathbb{R}^{K_{l,i}}} \mathbb{E}[|Z_{t_i} - \alpha\cdot p_{l,i}(X_{t_i})|^2] + C\mathbb{E}[|Z_{t_i} - Z_i|^2]$$

Plugging these estimates on $T_{1,i}^{\cdot,M}$ into $\mathcal{E}_{1,j}^M$ in (4.4.17) and (4.4.18) implies that that the upper bound of $\mathcal{E}(M)$ contains terms of the form

$$\Big(\sum_{j=0}^{N-1} \frac{\{\|Y_{t_j} - Y_j\|_2 + \|Z_{t_j} - Z_j\|_2\}\Delta_j}{(T-t_j)^{(1-\theta)/2}}\Big)^2 + \sum_{i=0}^{N-1}\Big(\sum_{j=i+1}^{N-1} \frac{\{\|Y_{t_j} - Y_j\|_2 + \|Z_{t_j} - Z_j\|_2\}\Delta_j}{(T-t_j)^{(1-\theta)/2}\sqrt{t_j - t_i}}\Big)^2\Delta_i.$$

In fact, it was a major task in the proof of Theorem 1.7.6 in Chapter 1 to show that these terms can be bounded above by $CN^{-2\theta_{\text{conv}}}$.

Let us assume that, for every $i \in \{0, \ldots, N-1\}$, there are measurable, deterministic functions $y_{t_i} : \mathbb{R}^d \to \mathbb{R}$ and $z_{t_i} : \mathbb{R}^d \to \mathbb{R}^q$ such that $Y_{t_i} = y_{t_i}(X_{t_i})$ and $Z_{t_i} = z_{t_i}(X_{t_i})$ almost surely. In the uniformly elliptic setting, one could take the functions given in [DG06, Theorem 2.1]. Moreover, we assume that, for some $n \geq 1$, the function $y_{t_i}(\cdot)$ is $(n+1)$-times differentiable, and

$$|\frac{\partial^r}{\partial x_{i_1} \cdots \partial x_{i_r}} y_{t_i}(x)| \leq C(T-t_i)^{(\theta_\Phi - r)/2} \qquad \forall x \in \mathbb{R}^d \tag{4.4.39}$$

for all $r \leq n+1$ and $i_1, \ldots, i_r \in \{1, \ldots, d\}$, and that the function $z_{t_i}(\cdot)$ is n-times differentiable, and

$$|\frac{\partial^r}{\partial x_{i_1} \cdots \partial x_{i_r}} z_{t_i}(x)| \leq C(T-t_i)^{(\theta_\Phi - (r+1))/2} \qquad \forall x \in \mathbb{R}^d \tag{4.4.40}$$

for all $r \leq n$ and $i_1, \ldots, i_r \in \{1, \ldots, d\}$. We remark that the differentiability assumptions are quite natural. We would like to mention (but will not use) the results of the recent paper [CD12], where gradient bounds of the form (4.4.39) and (4.4.40) are determined in the case were Φ is Lipschitz continuous (and even only measurable) and certain differentiability conditions on the coefficients b and σ and the driver f. The coefficients b and σ are time-inhomogeneous, but the uniformly elliptic condition is relaxed in that paper. The authors also do not consider Hölder continuous terminal conditions.

Using these additional differentiability assumptions, we calibrate the basis functions $p_{l,i}$. Let $\bar{R} > 0$ be a finite constant to be determined below. On $x \in [-\bar{R}, \bar{R}]^d$, we take $p_{0,i}(x)$ (resp. $p_{l,i}$ for l greater than 1) to be local polynomials of degree $n+1$ (resp. n), localized on $K_{0,i}/(n+1)$-hypercubes (resp. $K_{l,i}/n$) - denoted by $\{\mathcal{H}_{0,j} : j = 1, \ldots, K_{0,1}/(n+1)\}$ - of length $\delta_{y,i} = (n+1)^{1/d}\bar{R}/(K_{0,i})^{1/d}$ (resp. $\delta_{z,i} = n^{1/d}\bar{R}/(K_{l,i})^{1/d}$). We take $p_{l,i} \equiv 0$ outside $[-\bar{R}, \bar{R}]^d$ for all l. Using the Taylor expansion of y_{t_i} and the estimates (4.4.39) on each of the hyper cubes, it follows that

$$\begin{aligned}
&\inf_{\alpha \in \mathbb{R}^{K_{0,i}}} \mathbb{E}[|y_{t_i}(X_{t_i}) - \alpha \cdot p_{0,i}(X_{t_i})|^2] \\
&\qquad \leq |y_{t_i}|_\infty \mathbb{P}(|X_{t_i}|_\infty \geq \bar{R}) + (\delta_{y,i})^{2n} |\nabla_x^{(n+1)} y_{t_i}|_\infty^2 \sum_{j=1}^{K_{0,i}} \mathbb{P}(X_{t_i} \in \mathcal{H}_j) \\
&\qquad \leq C\mathbb{P}(|X_{t_i}|_\infty \geq \bar{R}) + C\frac{(\delta_{y,i})^{2(n+1)}}{(T-t_i)^{n+1-\theta_\Phi}}
\end{aligned}$$

Assuming X_{t_i} has exponential moments, it follows from Markov's inequality that $\bar{R} = C\ln(N)$ is sufficient for $\mathbb{P}(|X_{t_i}|_\infty \geq \bar{R}) \leq CN^{-2\theta_{\mathrm{conv}}}$. On the other hand, $\delta_{y,i} = N^{-\theta_{\mathrm{conv}}/(n+1)}(T-t_i)^{(n+1-\theta_\Phi)/2(n+1)}$ is sufficient for the other terms. $K_{0,i} = \left(\frac{CN^{\theta_{\mathrm{conv}}}}{(T-t_i)^{(n+1-\theta_\Phi)/2}}\right)^{\frac{d}{n+1}} \ln^d(N)$. Likewise, using estimates (4.4.40) to calibrate $\delta_{z,i} = N^{-\theta_{\mathrm{conv}}/n}(T-t_i)^{(n+1-\theta_\Phi)/(2n)}$, for all $l \in \{1, \ldots, q\}$, $K_{l,i} = \left(\frac{CN^{\theta_{\mathrm{conv}}}}{(T-t_i)^{(n+1-\theta_\Phi)/2}}\right)^{\frac{d}{n}} \ln^d(N)$ is required. It's clear that $K_{l,i} > K_{0,i}$.

Variance terms. We use Theorem 4.4.5 and Corollary 4.4.6 to calibrate M_i. Since $C_{\bar{z},i}^2 = C(T-t_i)^{1-(2\theta_c)\wedge\theta_\Phi}$, setting

$$M_i = \frac{CK_{l,i}N^{2\theta_{conv}}}{(T-t_i)^{1-\theta_c \wedge \frac{\theta_\Phi}{2}}} = \frac{CN^{(2+\frac{d}{n})\theta_{\mathrm{conv}}}}{(T-t_i)^{1+d\frac{n+1-\theta_\Phi}{2n} - (2\theta_c)\wedge\theta_\Phi}} \ln^d(N)$$

ensures that $T_{2,i}^{Z,M} \leq CN^{-2\theta_{conv}}$. Since $K_{l,i} > K_{0,i}$, this will dominate the requirement

$$M_i = CK_{0,i}N^{2\theta_{conv}}$$

to ensure that $T^{2,M}_{Y,i} \leq CN^{-2\theta_{\text{conv}}}$: if $(T-t_i) > 1$, then the terms $(T-t_i)^{(2\theta_c)\wedge\theta_\Phi - 1}$ are hidden in the constant C.

Large deviation events. Set $\varepsilon_{\dots} = N^{-2\theta_{conv}}$. Using Lemma 4.4.6, we see that we require

$$M_i = \frac{CN^{4\theta_{\text{conv}}}}{(T-t_i)^{2(1-(2\theta_c)\wedge\theta_\Phi)}}$$

to ensure that $\mathbb{P}^M_i(A^{Z,M}_i) \leq CN^{-2\theta_{\text{conv}}}$. This will be dominated by the requirements from the variance terms if $N^{\theta_{\text{conv}}d/n}(T-t_i)^{-d/2+(1-\theta_\Phi)/(2n)} > N^{2\theta_{\text{conv}}}(T-t_i)^{\theta_c\wedge\frac{\theta_\Phi}{2}-1}$, i.e. when the dimension is large. On the other hand, the requirements of $\mathbb{P}(B^{(Z,M)}_i)$ dominate those of the variance terms, because of the logarithmic terms. We take

$$M_i = \frac{CN^{(2+\frac{d}{n})\theta_{\text{conv}}}}{(T-t_i)^{1+d\frac{n+1-\theta_\Phi}{2n}-(2\theta_c)\wedge\theta_\Phi}} \ln^{2+d}(N)$$

Time-grids. In Chapter 1, we used the time grids $\{t_i = T - T(1-\frac{i}{N})^{\frac{1}{\theta_\pi}}\}$ for $\theta_\pi \in (0,1]$. This ensures that $\mathcal{E}(\pi) \leq CN^{-1}$, i.e. $\theta_{\text{conv}} = 0.5$, assuming that θ_Φ and θ_c are sufficiently large and θ_π sufficiently small; see Theorem 1.7.6. Generally speaking, we set $\theta_\pi = \delta\theta_\phi$ for some $\delta > 0$

Substituting this time-grid, and $\theta_{\text{conv}} = 0.5$, into the value of M_i, we obtain

$$M_i = \frac{CN^{1+\frac{d}{2n}+\frac{1}{\theta_\Phi\delta}+d\frac{n+1-\theta_\Phi}{2\theta_\Phi\delta n}-\frac{2\theta_c}{\theta_\Phi\delta}\wedge\frac{1}{\delta}}}{(N-i)^{\frac{1}{\theta_\Phi\delta}+d\frac{n+1-\theta_\Phi}{2\theta_\Phi\delta n}-\frac{2\theta_c}{\theta_\Phi\delta}\wedge\frac{1}{\delta}}} \ln^{2+d}(N). \tag{4.4.41}$$

We denote $\Psi := \frac{1}{\theta_\Phi\delta} + d\frac{n+1-\theta_\Phi}{2\theta_\Phi\delta n} - \frac{2\theta_c}{\theta_\Phi\delta} \wedge \frac{1}{\delta}$ for notational simplicity in the following computations.

Computational effort. The cost of computing the regression coefficients for the local polynomials costs $C\sum_{i=0}^{N-1} M_i$ flops [GVL96], whereas the cost sorting the simulations into the hypercubes is $\sum_{i=0}^{N-1} \ln(K_{l,i})M_i$ flops. The cost of generating the simulations is $N^2\sum_{i=0}^{N-1} M_i$, which is clearly dominant in this situation. Therefore, the complexity of the algorithm is dominated by

$$CN\sum_{i=0}^{N-1} M_i = CN^{2+\frac{d}{2n}+\Psi}\sum_{i=0}^{N-1}(N-i)^{-\Psi}\ln^{2+d}(N).$$

Since the summand is monotone increasing,

$$\sum_{i=0}^{N-1}(N-i)^{-\Psi} \leq 1 + \int_0^{N-1}(N-t)^{-\Psi}dt \leq \begin{cases} C & \text{if } \Psi > 1, \\ \ln(N) & \text{if } \Psi = 1, \\ CN^{1-\Psi} & \text{if } \Psi < 1. \end{cases}$$

Therefore, the complexity of this algorithm is dominated by

$$\mathcal{C} = CN^{2+\frac{d}{2n}+1\vee\Psi}\ln^{3+d}(N). \tag{4.4.42}$$

It is clear from the definition of Ψ that the complexity will become larger as the Hölder exponent gets smaller, therefore, the efficiency of the MWLS algorithm will deteriorate as the regularity of the terminal condition deteriorates.

In order to make a fair comparison to other adapted LSMDP of Chapter 3, we assume that the terminal condition is Lipschitz continuous; in this case, it is possible to take $\delta = 1$, and it follows that $\Psi \geq 1$ always. Note that $\Psi = d/2$. Therefore, the overall complexity of the scheme is

$$\mathcal{C} = CN^{2+d/(2n)+1\vee d/2}\log^{2+d}(N).$$

If we compare these results to the results of Proposition 3.5.2 of Chapter 3, the adapted LSMDP algorithm attains an complexity

$$\mathcal{C}_{aLSMDP} = C(N^{3+\frac{d}{2}+\frac{d}{2(n-1)}}\mathbf{1}_{d\geq 2} + CN^{4+\frac{d}{2(n-1)}}\mathbf{1}_{d=1})\log^{d+2}(N).$$

so the MWLS algorithm attains an improvement in the efficiency of order 1.

4.5 Appendix

4.5.1 Proof of Lemma 4.3.1

Consider the case $t_i = 0$ and $t_k = 1$. We set $\varphi(s) = (1-s)^{\alpha-1}s^{\beta-1}$ and we use the integral $\int_0^1 \varphi(s)ds$ (equivalent to the usual beta function with parameters (α, β)) to bound the sum. A simple but useful property is that φ is either monotone or unimodal (has increasing first derivative): thus,

$$(1-t_j)^{\alpha-1}t_j^{\beta-1}\Delta_j \leq R_\pi \int_{t_{j-1}}^{t_j} \varphi(s)ds + \int_{t_j}^{t_{j+1}} \varphi(s)ds.$$

Summing up over j and defining $B_{\alpha,\beta} = (1+R_\pi)\int_0^1 \varphi(s)ds$ concludes the proof for the simple case. For general t_i and t_k one can use the bounds on the simple case by rearranging the j-sum which is equal to

$$(t_k - t_i)^{\alpha+\beta-1}\sum_{j=i+1}^{k-1}(1-\frac{t_j - t_i}{t_k - t_i})^{\alpha-1}(\frac{t_j - t_i}{t_k - t_i})^{\beta-1}\frac{\Delta_j}{t_k - t_i} \leq B_{\alpha,\beta}(t_k - t_i)^{\alpha+\beta-1}.$$

□

4.5.2 Proof of Lemma 4.3.2

If $\alpha \geq \frac{1}{2}$, the result trivially holds with $\mathcal{C}_{(4.3.2a)} = 1$ and $\mathcal{C}_{(4.3.2b)} = C_u T^{\frac{(\alpha-1)}{2}}$.

Now, assume $\alpha < \frac{1}{2}$: if (4.3.1) holds, of course we also have

$$u_j \leq w_j + \sum_{l=j+1}^{N-1}\frac{w_l\Delta_l}{(T-t_l)^{\frac{1}{2}-\beta}(t_l - t_j)^{\frac{1}{2}-\alpha}} + C_u\sum_{l=j+1}^{N-1}\frac{u_l\Delta_l}{(T-t_l)^{\frac{1}{2}-\beta}(t_l - t_j)^{\frac{1}{2}-\alpha}}. \quad (4.5.1)$$

By substituting (4.5.1) into the last sum, and using Lemma 4.3.1 we observe

$$
\begin{aligned}
&\sum_{l=j+1}^{N-1} \frac{u_l \Delta_l}{(T-t_l)^{\frac{1}{2}-\beta}(t_l-t_j)^{\frac{1}{2}-\alpha}} \\
&\leq \sum_{l=j+1}^{N-1} \frac{w_l \Delta_l}{(T-t_l)^{\frac{1}{2}-\beta}(t_l-t_j)^{\frac{1}{2}-\alpha}} + \sum_{l=j+1}^{N-1} \frac{\sum_{r=l+1}^{N-1} \frac{w_r \Delta_r}{(T-t_r)^{\frac{1}{2}-\beta}(t_r-t_l)^{\frac{1}{2}-\alpha}} \Delta_l}{(T-t_l)^{\frac{1}{2}-\beta}(t_l-t_j)^{\frac{1}{2}-\alpha}} \\
&\quad + C_u \sum_{l=j+1}^{N-1} \frac{\sum_{r=l+1}^{N-1} \frac{u_r \Delta_r}{(T-t_r)^{\frac{1}{2}-\beta}(t_r-t_l)^{\frac{1}{2}-\alpha}} \Delta_l}{(T-t_l)^{\frac{1}{2}-\beta}(t_l-t_j)^{\frac{1}{2}-\alpha}} \\
&\leq \sum_{l=j+1}^{N-1} \frac{w_l \Delta_l}{(T-t_l)^{\frac{1}{2}-\beta}(t_l-t_j)^{\frac{1}{2}-\alpha}} + B_{\alpha+\beta,\frac{1}{2}+\alpha} \sum_{r=j+2}^{N-1} \frac{w_r \Delta_r}{(T-t_r)^{\frac{1}{2}-\beta}(t_r-t_j)^{\frac{1}{2}-2\alpha-\beta}} \\
&\quad + C_u B_{\alpha+\beta,\frac{1}{2}+\alpha} \sum_{r=j+2}^{N-1} \frac{u_r \Delta_r}{(T-t_r)^{\frac{1}{2}-\beta}(t_r-t_j)^{\frac{1}{2}-2\alpha-\beta}}.
\end{aligned}
$$

Substituting into (4.5.1), we observe that we have an equation of similar form to (4.5.1), except that, in the sum involving u, $\alpha \mapsto 2\alpha+\beta$ and $C_u \mapsto C_u B_{\alpha+\beta,\frac{1}{2}+\alpha}$, and, in the sum involvung w, $w \mapsto (1+C_u(1+T^{\alpha+\beta}B_{\alpha+\beta,\frac{1}{2}+\alpha}))w$.

After κ iterations of the previous step, we obtain $\alpha \mapsto 2^\kappa(\alpha+\beta)-\beta := \alpha_\kappa$. Hence, by iterating the above steps κ-times for κ such that $\alpha_\kappa \geq \frac{1}{2}$, i.e. $\kappa \geq \log_2\left(\frac{\frac{1}{2}+\beta}{\alpha+\beta}\right)$, we obtain the bound advertised in the Lemma statement. $\square$

4.5.3 Proof of Lemma 4.3.3

W.l.o.g we can assume for the proof that $\mathcal{C}_{(4.3.2a)} = 1$ in (4.3.2), up to replacing w by $w/\mathcal{C}_{(4.3.2a)}$. We first prove the case $\gamma = 1$. We need some extra notation. For some $\bar{\zeta} > 0$ (specified later), let

$$\zeta_s = \bar{\zeta} \int_0^s \frac{dr}{(T-r)^{\frac{1}{2}-\beta}} \leq \frac{2}{1+2\beta} \bar{\zeta} T^{(1+2\beta)/2}. \tag{4.5.2}$$

We write $\zeta_j = \zeta_{t_j}$ for brevity. Using (4.3.2) and switching the order of summation, we obtain

$$\begin{aligned}
&\sum_{j=i+1}^{N-1} \frac{u_j e^{\zeta_j}\Delta_j}{(T-t_j)^{\frac{1}{2}-\beta}} \\
&\le \sum_{j=i+1}^{N-1} \frac{w_j e^{\zeta_j}\Delta_j}{(T-t_j)^{\frac{1}{2}-\beta}} + \sum_{j=i+1}^{N-1} \frac{\sum_{l=j+1}^{N-1} \frac{w_l\Delta_l}{(T-t_l)^{\frac{1}{2}-\beta}(t_l-t_j)^{\frac{1}{2}-\alpha}} e^{\zeta_j}\Delta_j}{(T-t_j)^{\frac{1}{2}-\beta}} \\
&\quad + \mathcal{C}_{(4.3.2b)} \sum_{j=i+1}^{N-1} \frac{\sum_{l=j+1}^{N-1} \frac{u_l\Delta_l}{(T-t_l)^{\frac{1}{2}-\beta}} e^{\zeta_j}\Delta_j}{(T-t_j)^{\frac{1}{2}-\beta}} \\
&\le e^{\zeta_T} \sum_{j=i+1}^{N-1} \frac{w_j\Delta_j}{(T-t_j)^{\frac{1}{2}-\beta}} + e^{\zeta_T} B_{\alpha+\beta,1} \sum_{l=i+2}^{N-1} \frac{w_l\Delta_l}{(T-t_l)^{\frac{1}{2}-\beta}(t_l-t_i)^{-\alpha-\beta}} \\
&\quad + \mathcal{C}_{(4.3.2b)} \sum_{l=i+2}^{N-1} \frac{u_l\Delta_l}{(T-t_l)^{\frac{1}{2}-\beta}} \sum_{j=i+1}^{l-1} \frac{e^{\zeta_j}\Delta_j}{(T-t_j)^{\frac{1}{2}-\beta}} \\
&\le e^{\zeta_T}(1+B_{\alpha+\beta,1}T^{\alpha+\beta}) \sum_{l=i+1}^{N-1} \frac{w_l\Delta_l}{(T-t_l)^{\frac{1}{2}-\beta}} + \frac{1}{2}\sum_{l=i+1}^{N-1} \frac{u_l e^{\zeta_l}\Delta_l}{(T-t_l)^{\frac{1}{2}-\beta}}
\end{aligned}$$

where we have taken $\bar{\zeta} = 2\mathcal{C}_{(4.3.2b)}$ and used (since ζ is non-decreasing) so that because $\beta < \frac{1}{2}$ $\sum_{j=i+1}^{l-1} \frac{e^{\zeta_j}\Delta_j}{(T-t_j)^{\frac{1}{2}-\beta}} \le \int_{t_i}^{t_l} \frac{e^{\zeta_s}ds}{(T-s)^{\frac{1}{2}-\beta}} \le \frac{e^{\zeta_l}}{\bar{\zeta}} = \frac{e^{\zeta_l}}{2\mathcal{C}_{(4.3.2b)}}$. By subtracting the term with factor $\frac{1}{2}$, we obtain the result for $\gamma = 1$. Moreover, plugging the result into (4.3.2) gives

$$u_j \le \mathcal{C}_{(4.5.3)} w_j + \mathcal{C}_{(4.5.3)} \sum_{l=j+1}^{N-1} \frac{w_l\Delta_l}{(T-t_l)^{\frac{1}{2}-\beta}(t_l-t_j)^{\frac{1}{2}-\alpha}} + \mathcal{C}_{(4.5.3)} \sum_{l=j+1}^{N-1} \frac{w_l\Delta_l}{(T-t_l)^{\frac{1}{2}-\beta}} \tag{4.5.3}$$

for a constant $\mathcal{C}_{(4.5.3)} \ge 0$ depending only on $\mathcal{C}_{(4.3.2b)}$, T, α, and β.

Now for the general case $\gamma > 0$, observe that, for $\delta \ge 0$, we have

$$\sum_{j=i+1}^{N-1} \frac{\sum_{l=j+1}^{N-1} \frac{w_l\Delta_l}{(T-t_l)^{\frac{1}{2}-\beta}(t_l-t_j)^{\frac{1}{2}-\delta}}\Delta_j}{(T-t_j)^{\frac{1}{2}-\beta}(t_j-t_i)^{1-\gamma}} \le B_{\beta+\delta,\gamma} \sum_{l=i+2}^{N-1} \frac{w_l\Delta_l}{(T-t_l)^{\frac{1}{2}-\beta}(t_l-t_i)^{1-\beta-\delta-\gamma}}. \tag{4.5.4}$$

Thus, (4.5.3) becomes

$$\begin{aligned}
&\sum_{j=i+1}^{N-1} \frac{u_j \Delta_j}{(T-t_j)^{\frac{1}{2}-\beta}(t_j-t_i)^{1-\gamma}} \\
&\qquad \leq \mathcal{C}_{(4.5.3)} \sum_{j=i+1}^{N-1} \frac{w_j \Delta_j}{(T-t_j)^{\frac{1}{2}-\beta}(t_j-t_i)^{1-\gamma}} \\
&\qquad\quad + \mathcal{C}_{(4.5.3)} B_{\beta+\alpha,\gamma} \sum_{l=i+2}^{N-1} \frac{w_l \Delta_l}{(T-t_l)^{\frac{1}{2}-\beta}(t_l-t_i)^{1-\beta-\alpha-\gamma}} \\
&\qquad\quad + \mathcal{C}_{(4.5.3)} B_{\beta+\frac{1}{2},\gamma} \sum_{l=i+2}^{N-1} \frac{w_l \Delta_l}{(T-t_l)^{\frac{1}{2}-\beta}(t_l-t_i)^{\frac{1}{2}-\beta-\gamma}} \\
&\qquad \leq_c \mathcal{C}_{(4.5.3)}(1+B_{\beta+\alpha,\gamma}T^{\alpha+\beta}+B_{\beta+\frac{1}{2},\gamma}T^{\frac{1}{2}+\beta}) \sum_{j=i+1}^{N-1} \frac{w_j \Delta_j}{(T-t_j)^{\frac{1}{2}-\beta}(t_j-t_i)^{1-\gamma}}.
\end{aligned}$$

We are done. □

Lemma 4.5.1 (Conditional Hoeffding inequality)**.** *Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$, and suppose that $X_1, \ldots, X_n$ are $\mathbb{R}$-valued random variables that are "$\mathcal{G}$-conditionally independent" in the sense that, for any subset $\mathcal{Q} \subset \{1, \ldots, n\}$ and bounded measurable functions f_j, $j \in \mathcal{Q}$, the equality*

$$\mathbb{E}[\prod_{j \in \mathcal{Q}} f_j(X_j)|\mathcal{G}] = \prod_{j \in \mathcal{Q}} \mathbb{E}[f_j(x_j)|\mathcal{G}]$$

holds. Additionally, let $X_i \in [a_i, b_i]$ a.s. for some $-\infty < a_i \leq b_i < \infty$. Writing $S = X_1 + \ldots + X_n$, it follows that, for any $t \geq 0$,

$$\mathbb{P}(|S - \mathbb{E}[S|\mathcal{G}]| \geq t|\mathcal{G}) \leq 2\exp\Big(-\frac{2t^2}{\sum_{i=1}^n |b_i - a_i|^2}\Big)$$

The proof of Lemma 3.6.3 is analogous to [GKKW02, Lemma A.3]; one must only replace the independence with conditional independence and the use of the inequality $P(X > \epsilon) \leq e^{-t\epsilon}\mathbb{E}[e^{tX}]$ for all $t > 0$ with it's conditional version.

A Non-uniform time-grids

Lemma A.0.2. *The time grid $(t_k = T - T(1-k/N)^{1/\theta_\pi})_{0\le k\le N}$ with $\theta_\pi \in (0,1]$ satisfies*

$$C_\pi := \sup_{k<N} \frac{\Delta_k}{(T-t_k)^{1-\theta_L}} \le \frac{T^{\theta_L}}{\theta_\pi} \frac{1}{N^{1\wedge \frac{\theta_L}{\theta_\pi}}},$$
$$R_\pi := \sup_{0\le k\le N-2} \frac{\Delta_k}{\Delta_{k+1}} \le \frac{1}{\theta_\pi}\Big(1 \vee \big(\frac{1}{2\theta_\pi}\big)^{\frac{1}{\theta_\pi}-1}\Big),$$

where $\Delta_k = t_{k+1} - t_k$ and $\theta_L \in (0,1]$.

Proof. Set $1/\theta_\pi = \mu \ge 1$ and $g(x) = 1-(1-x)^\mu$: we have $t_k = Tg(k/N)$. Note that g is increasing and concave; thus we have

$$\frac{\Delta_k}{(T-t_k)^{1-\theta_L}} \le \frac{\frac{\mu T}{N}(1-k/N)^{\mu-1}}{T^{1-\theta_L}(1-k/N)^{\mu(1-\theta_L)}} = \frac{T^{\theta_L}}{\theta_\pi N}(1-k/N)^{\theta_L/\theta_\pi - 1}$$

and the bound on C_π follows by considering either $\theta_L \ge \theta_\pi$ or $\theta_L < \theta_\pi$.

Now, we study R_π. Since g is concave, we have $\Delta_{k-1} \ge \Delta_k \ge \cdots \ge \Delta_{N-1} = TN^{-\mu}$ and $\Delta_{k-1} \le \frac{\mu T}{N}(1-k/N)^{\mu-1}$. This gives a first upper bound for the n_0-last times $k = N-n_0, \ldots, N-1$ (with $n_0 \ge 1$):

$$\frac{\Delta_{k-1}}{\Delta_k} \le \frac{\Delta_{k-1}}{\Delta_{N-1}} \le \frac{\frac{\mu T}{N}(1-k/N)^{\mu-1}}{TN^{-\mu}} = \mu(N-k)^{\mu-1} \le \mu n_0^{\mu-1}. \tag{A.0.5}$$

We are now in a position to complete the upper bound on R_π.

- $\mu \in [1,2]$: we prove $\frac{\Delta_{k-1}}{\Delta_k} \le \mu$. For $k = N-1$, the inequality is true owing to (A.0.5). Now take $k < N-1$. Since g'' is non-increasing ($\mu \in [1,2]$), we have $\Delta_k \ge \frac{\mu T}{N}(1-k/N)^{\mu-1} + \frac{T}{2N^2}g''((k+1)/N)$, and we easily deduce

$$\frac{\Delta_{k-1}}{\Delta_k} \le \frac{\frac{\mu T}{N}(1-k/N)^{\mu-1}}{\frac{\mu T}{N}(1-k/N)^{\mu-1} - \frac{T}{2N^2}\mu(\mu-1)(1-(k+1)/N)^{\mu-2}} \le \frac{1}{1-\frac{(\mu-1)}{2}} \le \mu.$$

- $\mu \ge 2$: we prove $\frac{\Delta_{k-1}}{\Delta_k} \le \mu\big(\frac{\mu}{2}\big)^{\mu-1}$. Set $n_0 = \lfloor \frac{\mu}{2} \rfloor$: $n_0 \le \frac{\mu}{2} < n_0 + 1$. For $k \ge N - n_0$, the announced upper bound directly follows from (A.0.5). Now take $k \le N - n_0 - 1$ (which implies $N - k > \frac{\mu}{2}$): g'' being non-decreasing for $\mu \ge 2$, we have

$$\begin{aligned}\Delta_k \ge \frac{\mu T}{N}(1-k/N)^{\mu-1} + \frac{T}{2N^2}g''(k/N) &= \frac{\mu T}{N}(1-k/N)^{\mu-1}\big[1 - \frac{(\mu-1)}{2}(N-k)^{-1}\big] \\ &> \frac{\mu T}{N}(1-k/N)^{\mu-1}\frac{1}{\mu} \ge \Delta_{k-1}\frac{1}{\mu}.\end{aligned}$$

$\square$

References

[Avi09] R. Avikainen. On irregular functionals of SDEs and the Euler scheme. *Finance and Stochastics*, 13(3):381–401, 2009.

[BD07] C. Bender and R. Denk. A forward scheme for backward SDEs. *Stochastic Processes and their Applications*, 117(12):1793–1823, 2007.

[BDH$^+$03] Ph. Briand, B. Delyon, Y. Hu, E. Pardoux, and L. Stoica. L^p solutions of backward stochastic differential equations. *Stochastic Processes and their Applications*, 108(1):109–129, 2003.

[BG13] T. Benzineb and E. Gobet. Preliminary control variates to improve empirical regression methods. *Available on http://hal.archives-ouvertes.fr/hal-00802394*, 2013.

[BGM09] E. Benhamou, E. Gobet, and M. Miri. Smart expansion and fast calibration for jump diffusions. *Finance and Stochastics*, 13(4):563–589, 2009.

[BH08] P. Briand and Y. Hu. Quadratic BSDEs with convex generators and unbounded terminal conditions. *Probability Theory and Related Fields*, 141(3-4):543–567, 2008.

[BS12] C. Bender and J. Steiner. Least-squares Monte Carlo for BSDEs. In R. Carmona, P. Del Moral, P. Hu, and N. Oudjane, editors, *Numerical Methods in Finance*. Series: Springer Proceedings in Mathematics, Vol. 12, 2012.

[BT04] B. Bouchard and N. Touzi. Discrete time approximation and Monte Carlo simulation of backward stochastic differential equations. *Stochastic Processes and their Applications*, 111:175–206, 2004.

[CD12] D. Crisan and F. Delarue. Sharp derivative bounds for solutions of degenerate semi-linear partial differential equations. *Journal of Functional Analysis*, 263(10):3024–3101, 2012.

[DG06] F. Delarue and G. Guatteri. Weak existence and uniqueness for forward-backward SDEs. *Stochastic Processes and their Applications*, 116(12):1712–1742, 2006.

[EKPQ97] N. El Karoui, S. Peng, and M. C. Quenez. Backward stochastic differential equations in finance. *Mathematical Finance. An International Journal of Mathematics, Statistics and Financial Economics*, 7(1):1–71, 1997.

[FS04] H. Föllmer and A. Schied. *Stochastic finance: An introduction in discrete time*, volume 27 of *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, extended edition, 2004.

[GG11] S. Geiss and E. Gobet. Fractional smoothness and applications in finance. In *Advanced mathematical methods for finance*, pages 313–331. Springer, Heidelberg, 2011.

[GGG12] C. Geiss, S. Geiss, and E. Gobet. Generalized fractional smoothness and L_p-variation of BSDEs with non-Lipschitz terminal condition. *Stochastic Processes and their Applications*, 122(5):2078–2116, 2012.

[Gil08] M. B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.

[GKKW02] L. Györfi, M. Kohler, A. Krzyżak, and H. Walk. *A distribution-free theory of nonparametric regression.* Springer Series in Statistics. Springer-Verlag, New York, 2002.

[GL06] E. Gobet and J. P. Lemor. Numerical simulation of BSDEs using empirical regression methods: theory and practice. In *Proceedings of the Fifth Colloquium on BSDEs (29th May - 1st June 2005, Shangai) - Available on http://hal.archives-ouvertes.fr/hal-00291199/fr/*, 2006.

[GL07] E. Gobet and C. Labart. Error expansion for the discretization of backward stochastic differential equations. *Stochastic Processes and their Applications*, 117(7):803–829, 2007.

[GL10] E Gobet and C. Labart. Solving BSDE with adaptive control variate. *SIAM Journal on Numerical Analysis*, 48(1):257–277, 2010.

[GLW05] E. Gobet, J.P. Lemor, and X. Warin. A regression-based Monte Carlo method to solve backward stochastic differential equations. *Annals of Applied Probability*, 15(3):2172–2202, 2005.

[GM10] Emmanuel Gobet and Azmi Makhlouf. $\mathbf{L}_2$-time regularity of BSDEs with irregular terminal functions. *Stochastic Processes and their Applications*, 120(7):1105–1132, 2010.

[GVL96] G. H. Golub and C. F. Van Loan. *Matrix computations.* Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, third edition, 1996.

[HIM05] Y. Hu, P. Imkeller, and M. Müller. Utility maximization in incomplete markets. *The Annals of Applied Probability*, 15(3):1691–1712, 2005.

[IDR10] P. Imkeller and G. Dos Reis. Path regularity and explicit convergence rate for BSDE with truncated quadratic growth. *Stochastic Processes and their Applications*, 120(3):348–379, 2010.

[IDRZ10] Peter Imkeller, Gonçalo Dos Reis, and Jianing Zhang. Results on numerics for FBSDE with drivers of quadratic growth. In *Contemporary quantitative finance*, pages 159–182. Springer, Berlin, 2010.

[JS03] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2003.

[Kal02] O. Kallenberg. *Foundations of modern probability.* Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[KP92] P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.

[Kus03] S. Kusuoka. Malliavin calculus revisited. *The University of Tokyo. Journal of Mathematical Sciences*, 10(2):261–277, 2003.

[LGW06] J.P. Lemor, E. Gobet, and X. Warin. Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations. *Bernoulli*, 12(5):889–916, 2006.

[Mos10] T. Moseler. *A Picard-type iteration for backward stochastic differential equations: convergence and importance sampling*. PhD thesis, Fachbereich Mathematik und Statistik der Universität Konstanz, 2010.

[MZ02] J. Ma and J. Zhang. Representation theorems for backward stochastic differential equations. *The Annals of Applied Probability*, 12(4):1390–1418, 2002.

[Nua06] D. Nualart. *The Malliavin calculus and related topics*. Probability and its Applications (New York). Springer-Verlag, Berlin, second edition, 2006.

[Pol84] D. Pollard. *Convergence of stochastic processes*. Springer Series in Statistics. Springer-Verlag, New York, 1984.

[Pro05] Philip E. Protter. *Stochastic integration and differential equations*, volume 21 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2005. Second edition. Version 2.1, Corrected third printing.

[REK00] R. Rouge and N. El Karoui. Pricing via utility maximization and entropy. *Mathematical Finance. An International Journal of Mathematics, Statistics and Financial Economics*, 10(2):259–276, 2000. INFORMS Applied Probability Conference (Ulm, 1999).

[Ric10] A. Richou. *Étude théorique et numérique des équations différentielles stochastiques rétrogrades*. PhD thesis, Université Rennes 1, 2010.

[Ric11] A. Richou. Numerical simulation of BSDEs with drivers of quadratic growth. *The Annals of Applied Probability*, 21(5):1933–1964, 2011.

[RY99] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.

[Tib96] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B. Methodological*, 58(1):267–288, 1996.

[TT90] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Analysis and Applications*, 8-4:94–120, 1990.

[Zha04] J. Zhang. A numerical scheme for BSDEs. *The Annals of Applied Probability*, 14(1):459–488, 2004.

[Zha05] J. Zhang. Representation of solutions to BSDEs associated with a degenerate FSDE. *The Annals of Applied Probability*, 15(3):1798–1831, 2005.

Eidesstattliche Erklärung

Hiermit erkläre ich, daß ich die vorliegende Arbeit allein und nur unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe.

Berlin, July 16, 2013

Plamen Turkedjiev